VOL. XV JUNE, 1931 No. 3

The Journal of APPLIED PSYCHOLOGY

JAMES P. PORTER, EDITOR
Ohio University, Athens, Ohio
WITH A BOARD OF CO-OPERATING EDITORS

TABLE OF CONTENTS

DONALD G. PATERSON AND MILES A. TINKER. Studies of Typographical Factors Influencing Speed of Reading. VI. Black Type Versus White Type 241
GRACE HOLMES. The Relative Legibility of Black Print and White Print 248
R. H. WATERS. A Study of Customer Attitude 252
AUSTIN H. TURNEY. A Study of the Reliability of Judgments in Relation to the Certainty of the Judgments, to the Interval Between Judgments, and to the Number of Subjects Judged 259
JUNE E. DOWNEY. Inverse Trends in the Pressey X-O Test and the Downey Will-Temperament Test 273
H. METZNER AND E. M. BAILOR. Extent and Nature of Students' Knowledge of Psychology Before and After the First Course 279
WILLIAM E. WALTON AND BERLAH M. MORRISON. A Preliminary Study of the Affective Values of Colored Lights 294
HARRY W. CRANE. The Necessity of Psychometric Tests in the Study of Maladjustments 304
C. O. WEBER. The Aesthetics of Rectangles and Theories of Affection 310
G. R. SMITH. Mental Growth—A Criticism 319
NOTES AND NEWS 324
BOOK REVIEWS 327
NEW BOOKS AND PAMPHLETS 333

PUBLISHED BI-MONTHLY BY THE EDITOR
All communications should be addressed to the Editor,
Ohio University, Athens, Ohio
Printed by
THE SCIENCE PRESS PRINTING COMPANY
LIME AND GREEN STREETS,
LANCASTER, PA.

Entered as second-class matter June 8, 1894, at the post office at Baltimore, Maryland, under the act of March 3, 1879.
Application for transfer of the second-class mailing privilege to Lancaster, Pa., post office, pending.
Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917. Authorized December 21, 1922.

COPYRIGHT 1931, BY JAMES P. PORTER

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. P. Bode, Indiana University; W. V. Bingham, Personnel Research Federation, New York; June E. Downey, University of Wyoming; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; Joseph Jastrow, University of Wisconsin; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College; Henry M. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, Boston, Mass.; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; J. E. Wallace Wallin, Bureau of Special Education, 959 Harrison Avenue, Columbus, Ohio; G. M. Whipple, Dalton, Mass.; Helen Thompson Wooley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and offprints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by this Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $3.00 a year. Single numbers $1.00. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

STUDIES OF TYPOGRAPHICAL FACTORS INFLUENCING SPEED OF READING

VI. BLACK TYPE VERSUS WHITE TYPE[1]

DONALD G. PATERSON AND MILES A. TINKER
University of Minnesota

Although the question of the speed of reading black print on white paper as compared with white on black is of relatively small interest to the printing industry as a whole it does arise as an important problem within the limited field of advertising. Were it not for the practical difficulties of printing white on black it is probable that this arrangement would find wider use in advertising practice as a device for attracting attention because of its novelty. However, the advertiser to whom cost is a minor consideration wishes to know the extent to which he can utilize the contrast effects of white print on black background without at the same time sacrificing ease and speed of reading. For definite knowledge it is natural for the advertiser to turn to the psychology laboratory.

Unfortunately the information on this point contained in standard texts on the psychology of advertising is scanty and unsatisfactory. Scott, who was the first to recognize the problem, made a laboratory study of the apprehension of isolated letters exposed for a brief interval, utilizing one subject.[2]

[1] The writers gratefully acknowledge their indebtedness for aid in conducting these studies to the following: J. B. Lippincott Co., Philadelphia, for permission to reprint the Chapman-Cook Speed of Reading Tests, Forms A and B, for experimental purposes; the University of Minnesota Graduate School for research grant; and the psychology department of Northwestern University for permission to test advanced psychology students.

[2] W. D. Scott, The Theory of Advertising, Small, Maynard and Company, Boston, Mass., 1903, pp. 138–139.

Some letters appeared as black print on white background; some white on black; and the remainder white on gray. Thus his results bear more directly upon the question of attention value yielding little knowledge concerning legibility as affecting speed of reading. He emphasized the finding that "the black letters on a white background were seen oftener than the white type on a black background." Hollingworth after a brief discussion of possible reasons why black on white attracts more attention than white on black and without citing any experimental evidence states: "So far as acuity and legibility go there is no difference between the two arrangements."[3]

The only direct evidence available is given by Starch in his 1914 text.[4] He prepared two pieces of text alike in all respects except that one was printed with black on white and the other with white on dark gray. When these texts were read by forty persons the results showed a 42 per cent. advantage in favor of black on white. However, the findings are greatly reduced in significance because of the small number of subjects and failure to describe completely important features of the experimental conditions. In addition to these criticisms it must be kept in mind that the results throw no light upon the specific question before us since he employed white on dark gray rather than white on black.

Two other standard texts (Adams[5] and Poffenberger[6]) fail to mention this problem. Thus on this point, psychology, at present, has surprisingly little to offer the advertiser.

To determine the degree to which black on white differs from white on black so far as speed of reading is concerned,

[3] H. L. Hollingworth, Advertising and Selling, D. Appleton and Company, New York, 1920, pp. 76–78.

[4] D. Starch, Advertising, Scott, Foresman and Company, Chicago, 1914, pp. 189–190; also, Principles of Advertising, A. W. Shaw Company, New York, 1923, pp. 668–669.

[5] H. F. Adams, Advertising and its Mental Laws. The Macmillan Company, New York, 1916.

[6] A. T. Poffenberger, Psychology in Advertising. A. W. Shaw Company, Chicago, 1925.

the technique previously employed in the present series of studies on typography was utilized. The Chapman-Cook Speed of Reading Tests were printed in 10 point, Scotch Roman type, set solid with 19 pica line widths. Zinc etchings were made from this type and used in the three typographical arrangements as described in Table I. Forms A and B with

TABLE I

Black Print on White Background versus White Print on Black Background

Differences given are for the mean score on Form A minus the mean score on Form B in Group I. Forms A and B are identical typographically, showing black type on white background. In Group II Form A is printed as in Group I and Form B is printed with white type on a black background. Zinc etchings and white enamel paper stock were used for Forms A and B under both conditions. In the one case black ink was used on the raised *letter outlines;* in the other case black ink was used on the *raised background,* the letters appearing as white outlines. The three zinc etchings were cast from copy printed on white enamel paper stock. The type used was Scotch Roman, 10 point, 19 pica line widths, set solid. The mean score is the average number of paragraphs of 30 words each in the Chapman-Cook Speed of Reading Tests read in 1¾ minutes. In each sub-group, N = 140 advanced students in classes in Business Psychology and Abnormal Psychology, Northwestern University.

TEST GROUP	TEST FORM AND KIND OF PRINT	MEAN	$P.E._{dist.}$	$P.E._{M}$	DIFF. BETWEEN MEANS IN			$P.E._{diff.}$*	$\frac{D}{P.E._{diff.}}$
					Paragraphs	Words per second	Per cent.		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I	A, Bl. on Wh.	16.73	2.51	.21	0.59	0.17	3.7	.15	4.45
	B, Bl. on Wh.	16.14	2.12	.18					
II	A, Bl. on Wh.	17.29	3.07	.26	2.40	0.69	16.1	.15	16.14
	B, Wh. on Bl.	14.89	2.31	.20					

* $P.E._{diff.}$ values in column 9 were computed on the basis of the standard formula for determining the statistical significance of obtained differences for *correlated measures.* The correlations for the two sub-groups in this experiment were + 0.78 ± .02 for Group I; and + 0.82 ± .02 for Group II.

identical typography (black on white) were administered to a control group (Test Group I). The experimental group was subjected to Form A printed with black on white and Form

B printed with white on black. In gathering the data both the control and the experimental groups were tested under identical conditions. Each of the two test arrangements were distributed alternately to subjects in the class rooms. In this way students sitting in odd-numbered seats constituted the control group and those sitting in even-numbered seats composed the experimental group. The tests were administered to the Business Psychology and the Abnormal Psychology classes at Northwestern University on February 19, 1930. From the former class data for 59 students are included in our tabulations whereas from the latter class data for 221 students are utilized, making a total of 280 cases divided evenly into the control and the experimental groups. These college students were for the most part upper classmen.

The directions, appeal for cooperation, time-limits, and other conditions of testing were identical with the procedures employed in our previous work.[7] However, there was one important exception. In testing the large class in Abnormal Psychology the examiner, in his desire to hurry matters along so as not to encroach unduly on the time devoted to the regular work of the class, gave the order to begin work on Form B with a shorter pause between forms than usual. The result was that a number of the students were confused and could not locate Form B with the usual speed and dispatch. Since the time limit of $1\frac{3}{4}$ minutes is short this variation from the standard procedure undoubtedly tended to lower the average score for Form B. This point will be referred to again in our discussion of results.

Table I contains all of the computations essential for comparing speed of reading under the two different conditions. Test Group I serves as a control group. For this group, no significant difference between mean scores should be found since Forms A and B have been shown in our previous work

[7] See "Studies of Typographical Factors Influencing Speed of Reading," which have appeared in previous issues of J. of Appl. Psychol.

to be equal in difficulty when typography is identical as in this experiment. But a slight difference exists (0.59 paragraphs as shown in column 6) which is statistically significant ($\frac{D}{P.E._{diff.}} = 4.45$). This difference reveals the existence of the constant error described in the previous paragraph. This demonstration of a difference for Test Group I serves as a check on the results for Test Group II in which the attempt was made to test the effect of contrasting printing conditions. Thus the differences found in Test Group II must be discounted by approximately 0.59 of a paragraph, plus or minus $P.E._M$. As a matter of fact a marked difference was found for Test Group II which amounted to 2.40 paragraphs. Thus, when black letters appear on a white background, speed of reading is increased 16.1 per cent. as compared with speed of reading white letters on a black background. As previously noted a part of this difference must be discounted. When such a discount is made by subtracting, as a correction, 0.59 paragraphs from the obtained difference approximately a 10.5 per cent. advantage in speed of reading black on white as compared with white on black is disclosed.

DISCUSSION OF RESULTS

The results secured in this experiment are more in harmony with Starch's findings than with Hollingworth's assertion that so far as acuity and legibility are concerned there is no difference between the two printing arrangements. But Starch found a 42 per cent. differential in favor of black on white whereas the present study shows approximately a 10.5 per cent. advantage. The discrepancy is probably due to the fact that Starch compared black on white versus white on dark gray whereas we compared black on white versus white on black. If one of the main factors favoring fast reading is the degree of contrast between letters and background then white letters on a black background when contrast is at a maximum should prove much more effective than white letters

on a dark gray background where the contrast between letters and background is far less marked. In the present experiment, since both printing arrangements involve exactly the same amount of contrast it is not surprising that a less marked difference in effect on speed of reading is found.

The advertiser who contemplates utilizing white on black must decide whether or not the greater attention value of white on black through novelty will be sufficient to offset the disadvantage of a 10.5 per cent. difference in speed of reading. Perhaps the best compromise would be to adopt the rule that whenever white on black is employed as a device for attracting attention the amount of reading material should be reduced to a minimum so that the question of speed of reading the advertising text ceases to be a problem of any moment.

From a theoretical point of view, it is difficult to state just why black on white should be read much faster than white on black. Hollingworth attempted a theoretical explanation by asserting that we generally attend to the *objects* in our environment and by association identify them as dark spaces whereas we identify background with light spaces such as air, openings, sky, water, etc. Hence, when white letters appear on black, he states that they appear as *holes* in the *object*—that is, we tend to react to the background rather than to the letters. In addition, one might emphasize the rôle of specific reading habits as one of the important factors involved. Practically all of our reading throughout life involves reacting to black letters on a white background. Therefore, since our reading habits involve complex and delicate oculomotor adjustments, one would expect that the exact opposite of the normal reading situation would tend to upset those habits to a sufficient extent to retard the rate of reading. At all events, it would seem that these disturbances are concerned primarily with difficulties of word perception, since Holmes, who used isolated words cut from our printed texts, discovered that words printed in the usual manner can be perceived at a greater distance than words printed in white on black back-

ground (14.7 per cent. difference).[8] Further experimentation must be undertaken in order to discover the factors which are responsible for the observed differences reported by Holmes and by the present study.

SUMMARY

1. At the present time there is surprisingly little experimental evidence regarding speed of reading black print on white background as compared with white print on black background. This problem is of sufficient importance to the advertiser to warrant solution by experimental studies.

2. In order to secure further evidence on this problem, the writers compared the reading efficiency of college students when presented with standard text material printed black on white and white on black. In all, 280 subjects were employed, half of whom constituted a control group.

3. A "corrected" difference of approximately 10.5 per cent. was found in favor of the black on white printing arrangement.

4. From a practical advertising view-point, the results suggest that advertisers should follow the rule that whenever white on black is employed as a device for attracting attention the amount of reading text should be minimal so that the question of reading speed becomes relatively unimportant.

5. An adequate theoretical explanation of the obtained results awaits further experimentation.

[8] G. Holmes, The Relative Legibility of Black Print and White Print, J. of Appl. Psychol. (In press).

THE RELATIVE LEGIBILITY OF BLACK PRINT AND WHITE PRINT*

GRACE HOLMES

Psychological Laboratory, University of Minnesota

Kirschmann,[1] employing the method of indirect vision, discovered that both geometric forms and block capital letters were recognized at a greater distance when printed in white type on a black background than when printed in black type on a white background. There was a 16 per cent. advantage found for the white on black.

In reporting a somewhat similar experiment Scott[2] states that there is a tendency for black letters on white background to be more readily perceived than white letters on black. Starch[3] likewise found an advantage of 42 per cent. in favor of black type on white background when compared with white type on dark grey background. Speed of reading a given text was used as a measure of legibility in the latter experiment. The findings of both Scott and Starch are contrary to those of Kirschmann. Further comparisons of the comparative legibility of black and white print are needed.

The purpose of the present study is to determine, by the distance method, the relative legibility of words in black print

* The writer is deeply indebted to Dr. Miles A. Tinker under whose direction this study was conducted.

[1] A. Kirschmann, Uber die erkennbarkeit geometrischer figuren und Schriftzeichen im indirekten Sehen, Arch. f.d. ges. Psychol., 1908, 13, 352–388.

[2] W. D. Scott, The Theory of Advertising, Small, Maynard and Co., Boston, 1903, 138–139.

[3] D. Starch, Advertising, Scott, Foresman and Co., New York, 1914, 189–190; and Principles of Advertising, A. W. Shaw Co., Chicago, 1923, 668–669.

on a white background and words in white print on a black background. This method is approximately comparable to that of Kirschmann.

Twenty students of the University of Minnesota, twelve women and eight men, served as subjects in this experiment. The stimulus material consisted of 64 five-letter words printed in ten point Scotch Roman type in both black print on a white background and white print on a black background. Sixteen each of these words were mounted on black or white cardboard, 8½ x 11 inches, according as the background was black or white. The words were arranged in random order, four lines of four words each per card. In presenting the cards to the subject the black and white type were systematically varied to control practice and fatigue effects.

The apparatus for determining legibility by the distance method consisted of an extended bench, along which moved a sliding carriage holding the words to be read. One end of the bench was fastened to a small table, carrying a right-angled support with head rest attached. The reading material was uniformly illuminated with an intensity of 35 foot-candle per square foot. This apparatus and method are fully described by Tinker.[4]

The experiment was performed in a semi-darkened room. A sheet of words was put in the carriage, which was placed at the farther end of the bench, and the readings were begun. The subject was instructed to read the words at a uniform rate of about one per second, and to say "blank" for any word that could not be recognized. After completing one reading of the page, the stimulus material was moved to a position twenty centimeters nearer the subject's eyes. Then a second reading was completed and this procedure was continued until every word on the page had been read correctly. The experimenter recorded on a duplicate sheet the distance

[4] M. A. Tinker, The Relative Legibility of Modern and Old Style Numerals, J. Exper. Psychol., 1930, 13, 453–461.

from the subject, at which each word was read accurately. Frequent rest periods were allowed.

The results are presented in Table I. Individual as well as group averages are given to show the consistency of the

TABLE I

Comparative Legibility of Black Print on White Background with White Print on Black Background

SUBJECT NO.	BLACK ON WHITE		WHITE ON BLACK		DIFFERENCE BETWEEN MEANS
	Mean	$S.D._{dist.}$	Mean	$S.D._{dist.}$	
1.	157.7	34.5	133.6	45.1	24.2
2.	157.2	30.2	129.1	36.2	28.1
3.	176.8	30.8	143.2	32.8	33.6
4.	202.1	45.4	171.2	50.1	30.9
5.	156.5	31.1	121.2	28.2	35.3
6.	184.1	36.4	165.0	34.4	79.0
7.	160.6	36.4	120.0	34.4	40.6
8.	229.0	49.8	190.6	41.8	38.4
9.	165.6	28.4	143.1	34.2	22.5
10.	166.4	44.6	147.8	37.2	18.6
11.	191.4	42.4	145.0	43.0	46.4
12.	144.2	23.2	119.7	29.6	24.5
13.	139.3	22.8	116.2	31.8	23.1
14.	133.7	19.9	133.1	26.7	0.6
15.	157.1	31.7	134.6	36.4	22.5
16.	146.2	41.0	122.7	37.5	23.4
17.	117.8	32.3	99.7	21.1	18.1
18.	155.4	35.9	122.1	29.8	33.3
19.	148.7	44.2	130.2	40.6	18.5
20.	173.8	30.7	157.8	33.2	15.9

GROUP AVERAGE	Mean	$P.E._{dist.}$	$P.E._{Mean}$	Diff.	$P.E._{diff.}$	$\frac{D}{P.E._{diff.}}$
Black on White	160.3	43.2	.81	20.5	1.14	18.1
White on Black	139.7	42.1	.79			

trends. Inspection of the upper part of the table reveals a consistent advantage of black type on a white background for every individual. Each mean and sigma in this part of the table is based on sixty-four readings. The last column shows that the size of the differences between the means are all rather large, with a single exception (subject No. 14).

At the bottom of the table are given group means based on 1280 readings and the significance of the difference between the means. The average distance from the subject at which the black words on white background were perceived was 160.3 cm.; that for white on black 139.8 cm. The difference between averages is 20.5 cm. which represents an advantage of 14.7 per cent. for black print over white print. This difference divided by the P.E. of the difference equals 18.1, which indicates a very reliable difference.

No valid explanation of why these results are contrary to those of Kirschmann and in agreement with those of Scott and Starch is available at present. The subjects reported that in the case of white print, discrimination of details, by which one identifies letters was very difficult. The word appeared as a blur of white light with little or no definite form to many of the letters. Hollingworth[5] would explain the superior legibility of black print on white background on the basis of association. He says that we associate dark spaces with objects and light spaces with background. Therefore, white letters would tend to appear as merely holes in the object which is now the dark part. Investigations now in progress at Minnesota will throw further light on this problem.

In summary we may say, that when ten point type is employed the legibility of words printed in black type on a white background is 14.7 per cent. greater than white type on a black background.

[5] H. L. Hollingworth, Advertising and Selling, D. Appleton and Co., New York, 1920, 76–78.

A STUDY OF CUSTOMER ATTITUDE[1]

R. H. WATERS
University of Arkansas

This paper presents a brief study of the public attitude of merchants and salesmen as judged by their customers. The general question investigated may be stated as follows: Is the customer attitude of the successful merchant or storekeeper different from that of the unsuccessful? The report does not claim to be more than a preliminary investigation into a very complex field of social psychology. It may be granted that there are certain discoverable factors which make for the successful as opposed to the unsuccessful merchant. Numerous factors at once suggest themselves, prices, appearance of the house of business, cleanliness, ad infinitum, but there is also the possibility that certain personality characteristics as reflected in the manner of the merchant, his treatment of the customer, his readiness to accommodate, etc., make for his success or failure as a business man. It is the latter group of factors that are of greater interest for this study.

The method of the investigation may be described as follows: The Customer Attitude Blank[2] was employed. These blanks were given to adults of the writer's acquaintance (faculty members) and to a number of graduate students in the University of Arkansas. Each individual was asked to fill in two blanks, one for each member of 25 pairs of merchants who could be judged as successful and unsuccessful.

[1] Research paper No. 180, Journal Series, University of Arkansas. This study was done while the writer was acting as consultant for the Personal Analysis Bureau.

[2] This blank was prepared by Dr. E. G. Stoy, formerly of the Personal Analysis Bureau.

These merchants were known to the rater in each case, i. e., their judgments were to be based on their acquaintance with merchants or business men with whom they had had dealings. For unavoidable reasons certain individuals reported on only one member of such a pair. Thus, instead of having 75 pairs[3] rated we have ratings on 59 unsuccessful and 67 successful business men. This fact does not necessarily invalidate the data, since we may assume that the individual in ranking the successful merchant compares him with some sort of standard or estimate of the unsuccessful. For this reason we are including all the data collected.

A difficulty is encountered in the meaning of the terms "successful" and "unsuccessful." For the purpose of this investigation these two terms were given a narrow definition, viz.: The successful merchant is one with whom you like to do business or to continue doing business; the unsuccessful merchant is he with whom you would not like to do business, or with whom you would prefer to discontinue doing business. These definitions were included in the instructions given to the raters and none of them objected that the definitions were ambiguous.

The collected data are presented in the accompanying table. The first column gives the question as it appeared in the blank. The following three columns represent the percentages of the raters answering under each of the three categories; "Yes," "No," "Do not know," for the successful, the last three present the same data for the unsuccessful merchants. There were a few cases in which the questions were not answered in any way. These were few and scattering and were omitted.

Several comparisons are made possible by these data. If we arbitrarily use 75% as representative of a significant criterion, we may select those items which characterize the successful merchant. These are shown to be the following:

[3] The other pairs were contributed by Dr. E. F. Thelin, Syracuse, and Dr. D. A. Laird, Colgate.

TABLE 1

QUESTIONS	SUCCESSFUL			UNSUCCESSFUL		
	Yes	No	Do not know	Yes	No	Do not know
1. Is he a man of his word? Can you always depend upon him to keep his promises?	88	6	6	27	54	19
2. Is he always courteous? Even under trying conditions?	83	12	5	46	49	5
3. If he is courteous, is it reflected in the attitude of his employees?	84	6	10	40	49	11
4. Will he go out of his way to accommodate you on some minor matter?	85	9	6	31	60	9
5. Does he misrepresent things either knowingly or unknowingly?	12	79	9	54	29	17
6. When you do business with him do you have the feeling that he is genuinely interested in your needs?	91	9	0	7	92	2
7. Does he really know your needs?	59	20	21	14	76	10
8. Does he keep you well posted on market information?	54	34	12	12	71	17
9. Is he all business?	39	58	3	44	53	3
10. Does he get into long conversations about personal affairs?	9	84	8	36	56	8
11. Is he so anxious to please that he spoils the contact?	2	95	3	25	71	3
12. Does he try to load you up at an inopportune time?	3	97	0	39	52	9
13. Does he try to rush you into purchases?	6	94	0	52	48	0
14. Does he occasionally stop in just to say "hello"?	33	61	7	10	81	8
15. Does he get enough recreation of the sort which would result in more business?	34	21	45	12	40	49
16. Does he devote too much time to this sort of recreation?	3	48	48	12	38	50
17. Does he go about in a worried or preoccupied manner?	10	88	1	34	57	9
18. Does he seem too good to be true?—Can he enjoy a man's joke?	41	44	15	19	53	29
19. Does he carry a full line to meet your needs?	89	9	2	34	66	0
20. Can you depend upon him for uniform quality?	94	4	1	27	64	8
21. Are his prices fair?	94	3	3	47	46	7
22. Does he treat everybody equally well, both small customers and large?	88	5	8	27	54	19
23. Does his business have an air of stability?	97	3	0	34	60	5

TABLE 1—*Continued*

QUESTIONS	SUCCESSFUL			UNSUCCESSFUL		
	Yes	No	Do not know	Yes	No	Do not know
24. Does he seem to have his business well organized? Is it smooth running?	93	4	3	23	67	11
25. Does his credit department follow good practice?	65	8	27	21	34	45
26. What about adjustments—do you get full satisfaction?	88	2	11	24	53	22
27. Are his deliveries made with dispatch?	77	8	15	40	46	13
28. Is it hard to reach him on the telephone—do you have to go through a number of hands before you can talk with him?	10	81	9	14	72	14
29. When he is away are you as well taken care of as when he is on the job?	74	17	9	59	24	17
30. Does he malign others in similar business?	7	84	9	31	48	21

TABLE 2

Characteristics of the Successful Merchant

POSITIVE		NEGATIVE	
Item	% Yes	Item	% No
23 Stable business	97	12 Tries to load customer	97
20 Uniform quality	94	11 Overanxious to sell	95
21 Fair price	94	13 Tries to rush purchase	94
24 Well organized business	93	17 Worried, preoccupied air	88
6 Interest in customer	91	10 Long, personal conversation	84
19 Full line	89	30 Maligning competitors	84
22 Same treatment to all	88	28 Hard to reach	81
26 Satisfactory adjustments	88	5 Misrepresents goods	79
1 Honest	88		
4 Accommodation	85		
3 Courteous employees	84		
2 Courteous himself	83		
27 Deliveries with dispatch	77		

This table shows the item number, gives a brief description of each and the percentage "yes" and "no" answers for the positive and negative items respectively.

When we attempt to make the same type of characterization of the unsuccessful merchant we discover that none of the items yield a percentage of "yes" returns above 59%. This would show that our questionnaire was limited in the range of items called for. Likewise there are but three "no" items which reach or surpass the 75% mark. These are:

TABLE 3
Negative Items Characteristic of the Unsuccessful Merchant

Item	% No
6 Interest in customer	92
14 Stops in to say 'hello'	81
7 Knows customer's needs	76

Of these three items reference to the data of table 1 will show that item 6, "Interest in customer," is the only one which is found in the positive items of the successful merchant. Evidently our questionnaire is better adapted to a characterization of the successful merchant than of the unsuccessful.

A second type of comparison afforded by the data which is of interest is the selection of those items which discriminate most clearly between the two types of merchants. If we use as indicative of such discrimination those items which show a difference of 50% between the successful and unsuccessful business men the following items result: both positive and negative items are shown with the difference in the per cent returns for each. The difference in the case of both sets of items favors the successful merchant in each comparison.

The above tables bring out several facts which may be summarized as follows: 1. Customer attitude is not based entirely on the prices of the merchandise. In all comparisons we note that the items 23 and 24, "an apparently stable

TABLE 4

The Items which Most Clearly Discriminate

POSITIVE		NEGATIVE	
Item	diff.	Item	diff.
6 Interest in customer	84	6 Interest in customer	83
24 Organized business	70	24 Organized business	63
20 Uniform quality	67	20 Uniform quality	60
26 Satisfactory adjustments	64	14 Full line	57
23 Stable business	63	23 Stable business	57
1 Honest	61	4 Accommodation	51
22 Same treatment to all	61	5 Misrepresents goods	50
19 Full line	55		
4 Accommodation	54		

business," "a well-organized, smooth-running business," rank well up toward the top. These items might not be expected, a priori, to show this tendency. 2. Of the nine positive items in table 4 which favor the successful merchant, three of them concern the personal relationships between merchant and customer, items 6, 22, and 4; two refer to the stable and well-organized character of the business, items 24 and 23; two deal with the quality and quantity of goods, items 20 and 19; one item with the matter of accommodation, item 4; and one item with the honesty of the merchant, item 1. 3. Likewise the unsuccessful merchant is judged as such not on account of the prices he asks, but largely on the basis of his treatment of the customer as is emphasized by tables 2 and 3. 4. Again the negative items in table 4 which differentiate the unsuccessful from the successful include two items of personal relationship between merchant and customer, items 6 and 4; two items which refer to the stability and organization of his business, items 24 and 23; two items refer to the quality and quantity of the merchandise, items 20 and 14; and one with the misrepresentation of his goods, item 5.

In general, then, this study, although tentative in character, does emphasize the fact that a merchant's success or failure,

in the sense defined above, is conditioned not only by the price and quality of his goods but also by other less objective personality characteristics as well. These findings are in essential agreement with a recent paper by Phillips in which are reported the results of an experiment on shopping.[4] He finds also that the personality of the salesman is an important item which makes for or against success in selling.

[4] Phillips, D. E. Class shopping experiment. J. Appl. Psychol., 14, 1930, 42–53.

A STUDY OF THE RELIABILITY OF JUDGMENTS IN RELATION TO THE CERTAINTY OF THE JUDGMENTS, TO THE INTERVAL BETWEEN JUDGMENTS, AND TO THE NUMBER OF SUBJECTS JUDGED

AUSTIN H. TURNEY
University of Kansas

Notwithstanding the criticisms often directed against ratings of personality traits and of other variables, the value of judgments to psychology and education is very great. "For many types of psychological phenomena ratings are the only practical equivalents of objective measurements, and this applies especially though not exclusively to introspective or verbal report data."[1]

Because of their importance many efforts have been made to improve rating techniques and much has been written regarding rating and rating devices. "Owing to the immense importance of ratings in psychological experimentation, both pure and applied, constructive effort should be directed toward improving the means whereby ratings are obtained."[1]

Among the commonly recognized techniques used to improve ratings may be listed:

1. The training of raters.
2. Limiting the number of judgments to be placed in any one category thus forcing the distribution of judgments to approximate a normal curve.
3. The use of some form of "graphic" rating device.
4. The "man to man" comparison.
5. Precise descriptions for the various degrees of the trait to be rated.
6. Limiting the traits or variables to be rated to the smallest possible number.

[1] Freyd, Max. The Graphic Rating Scale. *Journal of Educational Psychology*, 14: 83–102, Feb., 1923.

7. Having the judges indicate how certain they feel about each judgment they make.

This paper presents the results of a study of the reliability of teachers' judgments of their pupils in relation to the degree of certainty of their judgments, in relation to the number of pupils rated, and to the interval between ratings. The study seeks to answer the following questions:

1. Does the judge who rates a few subjects do so more reliably than does one who rates a relatively large number?
2. Is the correlation between first and second judgments affected by the length of the interval between the two judgments?
3. Is a judge who indicates most of his judgments as very certain a more reliable rater than one who may feel less certain about most of his judgments?
4. Is the judge who rates a very few subjects more certain regarding his judgments than one who rates a relatively large number?

METHOD

Twenty teachers rated each of their pupils on nine traits twice during the school year. The interval between the first and second rating was on the average 61 days. The average number of pupils rated was 51.5. The traits upon which judgments were secured were:

Self-confidence
Industry
Leadership
Cooperativeness
Originality
Perseverance
Dependability
Ambition
Personal Attractiveness

The directions for rating were specific. Each teacher indicated his certainty of each judgment as shown in the following samples. The ratings were assigned values from 10 to 130 with 70 as the average. Sample directions and the sample ratings were printed on one side of the rating card. The nine

traits appeared on the other side in a manner similar to the sample except that the extremes were more carefully and fully described. Sample directions and the sample ratings are reproduced below:

SAMPLE DIRECTIONS

In each trait or characteristic named in the scale, compare the subject to be rated with the average child of the same age in the High School.

Try to make real distinctions. Do not rate a subject high or low on all traits simply because he is exceptional in some one or a few.

In rating more than one subject rate them all on one trait at a time. Then go through and rate them all on the next trait, and so on until the rating is completed.

In general it is not advisable to study too long over one trait. Make the best judgment you can and go to the next.

Do not omit any ratings.

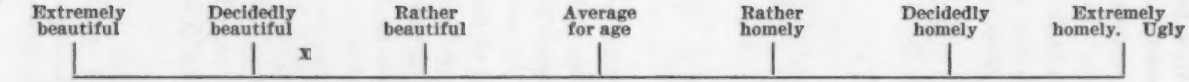
Below each line underline the word that best describes how certain you feel about your judgment. Do this for each trait that you rate.

The reliability of each judge's ratings was determined by calculating the Pearson product moment coefficient of correlation between first and second ratings. Since the intervening time was on the average two months these coefficients might be partly affected by changes in the amount of the trait shown by the pupils. That is to say Mary Smith might show more or less industry after the passage of two months' time. The coefficients obtained would therefore be less than they would have been had no changes in behavior occurred. It was desired, however, to remove all opportunity for memory to function in the second rating. Therefore the interval between first and second rating was in no case less than 45 days.

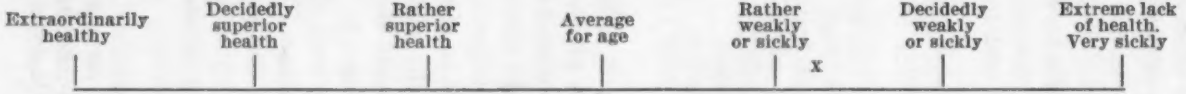
An index of certainty was calculated for each rater for each trait by averaging values assigned to the degrees of certainty as follows:

Very certain	1
Fairly certain	2
Rather uncertain	3
Very uncertain	4

SAMPLE RATINGS



My judgment on this trait was very certain, fairly certain, rather uncertain, very uncertain. (Underline.)



My judgment on this trait was very certain, fairly certain, rather uncertain, very uncertain. (Underline.)

Thus if Miss Smith rated 60 pupils in respect to the trait Self-confidence, and expressed herself as *very certain* regarding 10 pupils, *fairly certain* regarding 30 pupils, *rather uncertain* regarding 10 pupils, *very uncertain* regarding 10 pupils, her index of certainty would be $\frac{10X1 + 30X2 + 10X3 + 10X4}{60}$ or 2.33. This index was calculated for the first rating, for the second rating, and for both ratings combined.

It will be seen that the rating card used in this study is modeled after that used by Terman[2] in his study of gifted children, and that the index of certainty is the same as the mean certainty which is reported by him except that since the present study includes data from both a first and second judgment, an index of certainty for each rating and for the two combined was obtained. Apparently the mean certainty reported by Terman is from a single rating.

Our method is similar to that reported by Cady[3] in some respects but differs with regard to the method of determining reliability of the judgments. Cady correlated pairs of judgments of the same degree of certainty. If our interpretation is not in error, each member of a pair of judgments might be furnished by a different rater so long as both judgments were of the same degree of certainty as expressed by the judges and upon the same subjects. His data were apparently all from a single rating. In the present study the reliability of each judge is known from the correlation between first and second ratings. This study then compares the reliability of the judges with their own average degree of certainty. It will be seen that without a first and second rating it is not possible to tell whether the expression "very certain" means the same for all judges. Our data indicates that it does not. Without a first and second rating one could not tell whether a judge who usually expressed himself as *very* certain was any

[2] Terman, et al. Genetic Studies of Genius. Chapter 18.

[3] Cady, V. M. The Estimation of Juvenile Incorrigibility. Journal of Delinquency, Monograph No. 2, 1923.

more reliable than one who usually expressed himself as *fairly* certain. Our data indicates that the degree of certainty expressed by the judges is not a dependable index of actual reliability.

The reasons for using the method employed in this study are:

a. Each rater tended to express only one or two degrees of certainty. The majority of his judgments were indicated as of the same degree of certainty.
b. The degree of certainty varied from the first to the corresponding second rating of the same individual, in many cases.
c. The average degree of certainty expressed by any rater varied somewhat from the first to the second rating, but on the whole there was a close relation between them.
d. The raters showed definite individual differences in their degree of certainty.

To illustrate rater Mar (see table 1) rated 63 pupils. On the trait self-confidence for the first rating she expressed herself as very certain 43 times and rather certain 20 times. For the second rating she expressed herself as very certain 4 times and rather certain 59 times. There were only two pairs of judgments on the same pupils of the degree "very certain." Obviously this is not a large enough number to permit the comparison of the reliability of this rater's *very certain* judgments with her judgments of some other degree of certainty. To mix indiscriminately judgments made by different judges just because they are indicated as of the same certainty is confusing what some writers would call the validity with what others call reliability. As pointed out it does not follow according to our data that judgments indicated as of the same degree of certainty are of the same reliability.

The index of certainty used in this report represents the tendency of a given judge with respect to the certainty of his ratings, and since the judges differed consistently in this respect it makes possible the study of this factor in relation to others.

The data that follow are concerned with 19 judges who rated a total of 957 pupils, expressing in all 17,226 judgments.

RESULTS

Table 1[4] shows the reliability of each rater for each trait, the number of cases rated, and the average interval between the first and second rating. (Sample data only[5].) For convenience of interpretation there is also included in this table an average of the reliability coefficients for each trait and for each rater.

From table 1 it was found that

(a) Some traits are more difficult to rate reliably than others, notably Cooperativeness, or else students show greater changes in respect to this trait during the interval between ratings.

(b) There is considerable difference in the ability of individual instructors to rate reliably as shown by the range in the size of the coefficients (average) from .638 to .877.

(c) It is apparent that there is no perfect relationship between the size of the reliability coefficients and either the number of cases rated or the days interval between ratings. We shall consider these two relationships in greater detail.

In table 2 are shown the average reliability coefficients, the number of cases and the interval in days with the rank of each rater in each of these three variables.

The rank order coefficients of correlation between the size of the reliability coefficients and number of cases is $+.644$ (inferred r is $.658 \pm .087$);[6] between size of coefficient of correlation and interval in days is $+.116$ ($r = .126 \pm .149$).

[4] Turney, A. H., Factors other than Intelligence That Affect Success in the High School. Minneapolis. University of Minnesota Press. 1930. Table 19.

[5] In this and following tables only samples are shown. The complete manuscript containing full tables is on file at the library of the University of Kansas.

[6] Inferred values of r from rho taken from Table XX, Page 192 and P.E.'s from Table XVIII, p. 171 of Garrett's Statistics in Psychology and Education.

TABLE 1

Reliability Coefficients Obtained Between First and Second Ratings for Each Rater by Traits, Number of Cases and Interval Between Ratings (Sample data only)

	1. Self-confidence	2. Industry	3. Leadership	4. Cooperativeness	5. Originality	6. Perseverance	7. Dependability	8. Ambition	9. Personal Attractiveness	Average of Nine Coefficients	No. of Cases	Interval in Days
Rater												
Meg.	.826	.760	.653	.397	.343	.392	.611	.896	.863	.638	12	63
Nol.	.902	.935	.654	.779	.860	.327	.792	.826	.571	.738	15	61
Ing.	.902	.848	.883	.834	.898	.881	.898	.856	.893	.877	76	68
Dov.	.823	.470	.859	.814	.896	.825	.807	.806	.875	.798	101	61
Average of all judges	.803	.793	.758	.654	.754	.736	.770	.803	.810		51.5	61.6

TABLE 2

Average Reliability Coefficient, Number of Cases, and Interval in Days for Each Ratio. The Rank of Each Rater is Indicated for Each of These Variables (Sample data only)

Rater	Average r	Rank	N	Rank	Interval	Rank
Ing.	.877	1	76	3	68	3
Gar.	.876	2	65	5	60	11.5
Hur.	.782	8	21	17	51	18
Pat.	.780	9	47	10.5	67	4
Jas.	.683	17	56	8	80	2

It is, of course, apparent that there is no appreciable difference in reliability of judgment as between rater Ing. or Gar., or between Hur. and Pat., etc., but it is true that so far as this group is concerned rating fewer cases did not increase reliability, but that instead there is a rather decided tendency for the reliability to increase as the number of cases increased.

It is also apparent that reliability was not increased as a result of a shorter interval between the first and second rating.

The data indicate that the interval had little or nothing to do with the reliability of judgments even though the range in the interval between first and second rating was forty-four days.

The relation between the reliability of judgments (or judges) and the certainty of judgments (or judges) was studied in the following way: It was assumed that the average of the reliability coefficients obtained from each rater for the nine traits (see table 1) was an index of a general ability to rate reliably in comparison with any other of the judges. The relation between these average coefficients and the index of certainty for the first judgments, for the second judgments, and for the first and second judgments combined was studied.

Table 3 shows for each judge the reliability coefficients, the index of certainty for the first ratings, for the second ratings, and for the combined ratings.

TABLE 3

Average Reliability Coefficients, Index of Certainty for First Rating, for Second Rating, and for Both Ratings Combined, by Raters. The Rank of Each Rater in Each Variable is Also Shown (Sample data)

Rater	Average r	Rank	Avg. Index of Certainty 1st Rating	Rank	Avg. Index of Certainty 2nd Rating	Rank	Avg. Index of Certainty both ratings combined*	Rank
Gar.	.876	2	2.206	17	2.176	18	2.191	18.5
Sto.	.796	7	1.063	1	1.066	1	1.065	1
Sho.	.711	15.5	1.412	3	1.916	10	1.664	6

* This index is not an average of first and second index but was obtained from the original values.

The data of table 3 will show that there is not a close relation between the reliability of the judges and their own feeling of certainty. Rater Sto was always confident of his judgments. His index 1.063 shows that practically all of his judgments were indicated as very certain. No other rater was so sure of himself, yet six other raters were actually more reliable. Rater Gar. was decidedly less sure of his judgments. His index 2.206 shows that he indicated most of his judgments as rather certain with a number as rather uncertain. Yet he is second in the group in ability to rate reliably. Sho. was third in her feeling of confidence with regard to her judgments, but 15th in actual ability to judge reliably. It is evident that judgments which are indicated as very certain by one judge may be no better than judgments indicated as rather certain by another.

For convenience of interpretation the rank difference correlation between the size of the reliability coefficients and the indexes of certainty were calculated. The coefficients are:

	rho	r	p.e.
Between size of average coefficient and first index of certainty	+.374	.385	±.14

Between size of average coefficient and second index of certainty	+.487	.508	± .11
Between size of average coefficient and combined index of certainty	+.412	.426	± .13

It will be seen that there is a positive relation though not a marked one between the average coefficients of reliability and the indexes of certainty. There is some tendency then though not a pronounced one for the rater who is most confident of his judgments to rate more reliably than the rater who is less confident.

Since some objection might be raised regarding the use of the average of the coefficients of reliability as an index of the ability of each rater to rate reliably, further analysis of the data was made. The relation between the reliability coefficients and the index of certainty for each trait was studied. Table 4 shows sample data for the trait Self-confidence.

TABLE 4

Reliability Coefficient and Index of Certainty (for first and second judgments combined) for Each Rater for the Trait Self-confidence. The Rank of Each Rater is Also Shown (Sample data)

Rater	r	Rank	Index of Certainty	Rank
Ing.	.902	1.5	1.65	7
Dov.	.823	12	1.12	1
Pat.	.729	15	2.04	16

Rho = .218 (r = .229 ± .145)

The data for the other eight traits were treated as in table 4. These data are omitted to conserve space. The rho and inferred r for each trait are as follows:

	rho	r
I. Self-confidence (table 4)	.218	.229 ± .145
II. Industry	.098	.105 ± .149
III. Leadership	.779	.794 ± .054
IV. Cooperativeness	.253	.261 ± .141

V. Originality	.427	.447 ± .119
VI. Perseverance	.548	.568 ± .104
VII. Dependability	.475	.492 ± .113
VIII. Ambition	.291	.303 ± .137
IX. Personal Attractiveness	.355	.369 ± .132

These correlations bear out the data in table 3. The reliability of these coefficients is, of course, affected by the small number of cases (19 raters). However, they suggest that there is a positive but low and inconsistent relationship between the degree of certainty of the raters and their reliability. In a few traits the relationship is so small as to be of no importance. It is of interest to note that the size of the reliability coefficients and the certainty of judgments correlated highest in the case of the trait Leadership, although the judges as a whole actually rated five other traits more reliably than they did the trait Leadership. This is additional evidence that the degree of certainty expressed by judges is of no great value in indicating the reliability of their judgments.

The final step in this study was to discover if possible whether or not the degree of certainty of the judgments was affected by the number of cases rated. For this purpose the index of certainty for the first rating, the second rating and the combined ratings was studied in relation to the number of cases. Table 5 presents sample data.

TABLE 5

Average Index of Certainty for First Judgment; Second Judgment; First and Second Judgment; and the Number of Cases. The Rank of Each Judge is Shown. (The averages are averages from nine traits. Sample data only)

Rater	Avg. index first rating	Rank	Avg. index second rating	Rank	Average combined rating	Rank	N	Rank
Gar.	2.206	17	2.176	18	2.191	18.5	65	3
Sun.	1.498	4	1.598	4	1.598	4	51	9
Sho.	1.412	3	1.916	10	1.664	6	39	13

The rank order correlations with the inferred values of r for table 5 are:

	rho	r
Between number of cases and index of certainty first rating	.220	.229 ± .145
Between number of cases and index of certainty of second rating	.153	.157 ± .146
Between number of cases and combined index	.166	.167 ± .146

Apparently the feeling of certainty which the judges felt was not affected by the number of cases which each rated. Thus Dov., who rated the largest number of pupils, was second in rank in respect to degree of certainty.

SUMMARY OF FINDINGS

In the study it was found that

1. Ratings are not more reliable when a few cases are rated. Increasing the number of cases within reasonable limits does not decrease the reliability of judgments. A judge rating 100 subjects may do so more reliably than one rating only 12.

2. The interval between first and second ratings does not seem to affect the reliability of judgments greatly. A judge making his second ratings 90 days after the first ratings may be more reliable than one making his second rating 60 days after the first.

3. The degree of certainty has a low but positive and inconsistent relationship with reliability of judgment.

4. A judge may be very certain of his judgments and yet be less reliable than one who is considerably less certain.

5. The degree of certainty felt by judges is not greater when the number of cases is small. A judge rating 100 subjects may feel more certain than one rating only 40 subjects.

6. A close agreement of high correlation between the degree of certainty and obtained reliability on a given trait does not

mean high obtained reliability. A trait may be rated with relatively less reliability than another even though the degree of certainty correlates more highly with the obtained reliability for the first than for the second trait.

7. It is at least open to question whether or not expressing degree of certainty has any practical value. This can only be determined by a controlled experiment. Rater Gol. (see table 1) who expressed no certainty stood fourth in twenty in respect to reliability of judgment. She declared that to indicate certainty affected her judgments adversely.

CONCLUSIONS

If our data *are* valid it would appear that the ability to rate others reliably—in this case for teachers to rate pupils—is one that shows rather marked individual differences. Within reasonable limits it is not greatly affected by the number of cases or the length of the interval between ratings.

The reliability of judgments should be determined by correlating a first and second judgment. To depend upon the degree of certainty expressed by the rater as a criterion of reliability might mean discarding reliable judgments made by a less confident but more able judge.

INVERSE TRENDS IN THE PRESSEY X–O TEST AND THE DOWNEY WILL-TEMPERAMENT TEST

JUNE E. DOWNEY
University of Wyoming

In a recent article in this Journal[1] a comparison was made of the three following tests: the Downey Group Will-Temperament Test, the Pressey Cross-Out Tests and the Brotemarkle Comparison Tests. The three of them are described as attempts to analyze the complex emotional make-up of the individual with the implication that high positive correlations should be found between the total scores yielded by the three tests.

Attention should be called to the fact that the Will-Temperament Test is not correctly described as a test of the emotions. The aim in organizing it was to get some sort of expression of an individual's *dynamic* pattern. The specific tests carry with them no direct implication of emotional make-up. One, therefore, finds great difficulty in understanding why any investigator should expect to find a significant correlation between the Total Affectivity Score on the Pressey X–O Test and the total score on the Will-Temperament Test. The authors of the article under consideration make the following curious inquiry: "Is Downey justified in stating in her Test Manual that a high total score on her test indicates a strong personality while a low total score represents a weak personality? If so, why is it that no significant correlations are found between her total scores and those of the other tests?"[2] Since Total Affectivity may be interpreted as emotional spread it is difficult to see why a

[1] Gorham, D. R. and Brotemarkle, R. A., "Challenging Three Standardized Emotional Tests for Validity and Employability," 1929, 13, 554–588.

[2] *Loc. cit.*, p. 558 and 564.

high score on it would necessarily characterize a strong personality if the latter be interpreted to imply the possession of aggressiveness, speed qualities and energy. One might, perhaps, more logically conjecture that a negative relationship exists between affectivity spread and dynamic qualities. The authors of the article under consideration actually cite a correlation of $-.34$ for the Total Affectivity score on the Pressey Test and the total score on the Downey Group Will-Temperament Test—a most interesting outcome if it should be confirmed by work with a large group.

Here it should be emphasized that the Group Will-Temperament Test was used in the investigation under consideration, although, strangely enough, the names of the specific traits were taken from those employed in the Individual Will-Temperament Test. Since the two forms are not thoroughly comparable this procedure introduces some error in the checking up of results. The Individual Will-Temperament Test is probably a much more effective instrument than the Group Test and should, if possible, be used in clinical work. The conditions that operate when a group is tested influence to a considerable degree the scores on certain specific items.

To repeat, I can see no reason to anticipate a positive correlation between the Total Affectivity score on the Pressey X–O Test and the total score on the Will-Temperament Test. At most one might anticipate inverse relationship between certain trends revealed by these tests. Since it has proved a very difficult matter to determine what methods should be used to check the validity of tests of emotional, temperamental and volitional capacities, any suggestions of possible procedure merit consideration.

The Pressey Test was in part formulated on the basis of familiarity with symptoms exhibited by psychopathic patients. Test 1 and Test 4 are planned to evoke if possible certain emotional reactions. Test 4 may make manifest any of the five following trends: paranoid, neurotic, self-conscious or regressive, melancholic, and hyperchondriacal. The motives operating

in the choice of words for the test become evident if one follows through the series that has to do with a paranoid trend; for example, injustice, poison, enemies, boss, conspiracy, neighbors, insult, influence, etc. The implications seem fairly clear to individuals who have had experience in psychopathic hospitals.

One of the most promising attempts made to validate the various tests of the Will-Temperament Series consisted in giving them to psychopathic or psychotic subjects with the anticipation that certain types of patients would score high or low on certain specific tests.[3] Manic patients, for example, were expected to score high on motor impulsion. The few records at my disposal gave the expected reaction to a surprising degree. Melancholic patients were expected to score low on the same tests, and did so in the few cases tested. Other significant results were as follows: dementia praecox patients gave great retardation in speed, excessive load, low scores on motor inhibition and coordination of impulses, but *relatively* high scores for motor impulsion and reaction to contradiction.

In general, reactions to the tests for motor impulsion and motor inhibition were the most impressive in the case of hospital patients. In a sense these items are not so much tests of specific traits as indications of certain form-qualities that characterize various patterned reactions. The only patients scoring excessively low on motor impulsion were those suffering from depression; on the other hand, the scores on motor inhibition were uniformly low and there was also a uniformly low score on coordination of impulses.

Unfortunately the work on psychotic subjects has been very limited in extent. Since the publication of the book, "The Will-Temperament and its Testing," only twenty-six records on the Individual Form of the test have been added to the original series of psychotic subjects, and these constitute a scattering group of various cases. It may, however, be said that the profiles obtained from this second group are very like those obtained earlier. Again, motor impulsion scored much

[3] Downey, J. E., "The Will-Temperament and its Testing," 204f.

higher than motor inhibition. No patient scored over two for motor inhibition, while the average on motor impulsion was 4.84, the range being from 1.75 to 10. Again, there was the same difficulty apparent in handling coordination of impulses. Only two patients scored above one on this test. Reaction to contradiction ran higher than reaction to opposition—a characteristic previously noted for dementia praecox and psychopathic patients and one, I believe, of particular significance.

Since there are difficulties in the way of obtaining a large number of Will-Temperament records from mental patients, it would be highly desirable to determine what might be done by selection of cases for study by help of the Pressey X–O Tests. If definite trends are exhibited in a reaction to the Pressey Test, it might be possible to institute comparison with certain specific records on the Will-Temperament tests. The particular comparisons I have in mind are, of course, Will-Temperament scores on motor inhibition and motor impulsion, and Pressey scores on depressive or melancholic trends.

Difficulties in the way of accurate investigation are, however, very great. McGeoch and Whitely have shown, for example, that there is a fairly high self-correlation for the Pressey X–O Test when the two givings of the test are separated "by a 48 hour interval, with a tendency to a decrease in the consistency of measurement, both quantitatively and qualitatively, when 45 days and 90 days separate the testings."[4] Therefore, the Pressey Test and the Will-Temperament Test should be given with only a short interval of time between. Moreover, the sexes should be kept separate since in the Will-Temperament Test there are slight indications of sex-differences notably on the score for motor inhibition. Furthermore, it would be well if there could be a wide range selection of subjects rather than the close one that prevails in most college investigations.

[4] McGeoch, John A. and Whitely, P. L., "The Reliability of the Pressey X–O Tests for Investigating the Emotions," Pedagog. Seminary & Jour. of Genetic Psy., 1927, 34, 255–270.

Possibly a similar result may be found for the Will-Temperament Test.

I am reporting below a few correlations between items on the Pressey and the Will-Temperament Tests. The group tested consisted of twenty-eight students in psychology classes. The conditions under which they were obtained were far from ideal. A miscellaneous collection of material was used; the group is a mixed one, both men and women; and a number of the group were left-handed or shift-overs—a matter of some importance when a motor activity is used in scoring. Correlations were obtained by the method of rank-differences:

$$\rho = 1 - \frac{6\Sigma d^2}{\eta(\eta^2 - 1)}.$$

With these limitations in mind, table 1 may be studied as a

TABLE 1

Correlations between the results of parts of the Pressey X–O Test and the results of the Will-Temperament Test

1. Total Will-Temperament Score with Total Affectivity Score, Pressey X–O	+.294	± .121
2. Total Will-Temperament Score with Total Idiosyncrasy Score, Pressey X–O	+.118	± .131
3. Speed of Movement, Freedom from Load, Flexibility, Will-Temperament Test, with Total Affectivity Score, Pressey X–O	+.128	± .130
4. Speeded Movement, Will-Temperament Test, with Total Affectivity Score, Pressey X–O	+.05	± .133
5. Motor Inhibition, Will-Temperament, with Neurotic Responses, Test 4, Pressey X–O	+.068	± .132
6. Motor Impulsion, Will-Temperament Test, with Total self-conscious, melancholic, hypochondriacal responses, Test 4, Pressey X–O	+.18	± .128
7. Motor Inhibition, Will-Temperament Test, with Total self-conscious, melancholic, hypochondriacal responses, Test 4, Pressey X–O	+.06	± .132

summary of results. A low positive correlation was found between the total Will-Temperament score (Individual Form) and the Total Affectivity score on the Pressey X–O Test. When the Total Idiosyncrasy score was used instead of the Total Affectivity score the correlational coefficient was considerably reduced.

Since the Will-Temperament score includes items that frequently cancel one another, it seemed worthwhile getting the correlation between the Total Affectivity score and the speed items of the Will-Temperament Test where a negative correlation might be found. Again, the correlational coefficient was very low with a further reduction when a correlation was obtained for the Total Affectivity score and speeded movement.[5] Three other correlations were obtained. The score on motor impulsion was correlated with the total self-conscious, melancholic and hypochondriacal responses on the Pressey X–O, Test 4, and motor inhibition with the same three and also with the score on neurotic responses, Test 4, Pressey. The low coefficients and high probable errors indicate lack of positive correlation and suggest the value of carrying the investigation farther. Negative correlations would certainly go far to indicate the validity of both tests.

The above discussion is purely suggestive. Outstanding cases exhibiting different trends on the Pressey Test might be used in connection with the Will-Temperament Test in a most illuminating way.[6] In the article by Gorham and Brotemarkle cited at the opening of the discussion, the value of use of these tests in clinical work is illustrated in a most interesting way. What I am suggesting is the possible quantitative statement of relationships.

SUMMARY AND CONCLUSIONS

A possible method of checking the validity of specific items of the Will-Temperament and Pressey Cross-Out Tests is suggested by the fact that the two tests measure, to a certain extent, inversely related trends. The present investigation seems to indicate the possibility of negative correlations, under ideal conditions of testing between certain parts of the two tests.

[5] Speeded movement has been found to be a more reliable item than natural speed and should be substituted for the latter in Will-Temperament investigations.

[6] I am, however, inclined to believe that highly introverted persons react defensively to the Pressey Test. Allport has made the same suggestion, "Social Psychology," p. 133.

EXTENT AND NATURE OF STUDENTS' KNOWLEDGE OF PSYCHOLOGY BEFORE AND AFTER THE FIRST COURSE

H. MELTZER

Psychiatric Clinic and Washington University, St. Louis

AND

E. M. BAILOR

Dartmouth College

"Till the assumption is banished that stimulus to knowing and the object of knowledge are the same thing, the analysis of knowledge and truth will be confused."[1] Apparently James Harvey Robinson is of the opinion that this warning, dated 1922, has not been heeded in the field of higher education, for he thinks "a college is a place where there is much teaching and no learning."[2] Were the relationship between the stimulus to knowing and the object of knowledge closer, that between teaching and learning also would be nearer. Dealing with a narrower and more tangible aspect of the larger problem of the relation of teaching and learning Dean Worcester, in an article[3] on tautological teaching in higher education, reports findings which demonstrate the falsity of a particularized aspect of the assumption condemned by Dewey. The assumption is that all students who are to begin a course know nothing about the subject matter of the course. In his study Worcester concerned himself only with the extent of knowledge possessed by students upon entering and after having a course in Educational Psychology. His study an-

[1] Dewey, John, *Knowledge and Speech Reactions.* The Journal of Philosophy, October 12, 1922, Vol. XIX, No. 21.

[2] Quoted in Martin, E. D., *The Meaning of a Liberal Education.*

[3] Worcester, Dean A., *Twice Told Tales or Tautological Teaching in Higher Education.* Educational Administration and Supervision, February, 1926; vol. XII, No. 2, pp. 73–79.

swered the question "how much?" It did not touch on the more discriminatory but also tangible aspect of the problem, the nature as well as the extent of what is known. In the present study an attempt is made to do so by an analysis of (a) knowledge possessed and not possessed by students before and after taking the first course in psychology; (b) knowledge students learn and fail to learn; and (c) the relationships which may exist between knowledge gained during a first course and intelligence, socio-economic status, and age.

The number of cases studied is comparatively small and the results, therefore, cannot serve as a basis for generalizations about the relation of teaching and learning. The chief purpose of this study is not to prove conclusively any one fact, but rather to demonstrate a more prehensive approach to the study of such problems by reporting the findings from the one group in a described setting. The results will be more intelligible if, before they are reported, the measuring instruments and procedures used in obtaining them are first presented.

I. DESCRIPTION OF PROCEDURES AND MEASURING INSTRUMENTS

A. *General Procedure.*

The institution in which this experiment was made is a land-grant college, with an enrollment of 3,700, located on the Pacific Coast. On their first appearance in class a final examination was given to the group of 56 students who were registered in their first course in general psychology. At the end of the course the same final examination was given, but inasmuch as eleven registered for the course after the first day there were 67 who took the second examination. Scores on intelligence, socio-economic status, and other information used were obtained during the quarter.

B. *Measuring Instruments.*

1. Knowledge of psychology. For measuring knowledge of psychology before and after the course objective tests were used exclusively. The examination was composed of 100 elements set in three different forms: true-false, matching, and

multiple-choice. The tests were scored on the basis of the numbers right minus wrong for true-false elements and the number right for matching and multiple-choice ones. The total score was, of course, the total of the scores on the three different parts of the examination.

2. Intelligence. Intelligence was determined by the use of the Otis Self-Administering Test of Mental Ability, Higher Examination; Form A.

3. Socio-economic Status. Socio-economic status was measured by the Chapman-Sims Scale according to the standardization reported by the authors.[4] Socio-economic facts considered are: attendance of mother and father in high school; presence or absence in the home of: 125 or more books; a piano; two or more magazines; a telephone; an auto, furnace heat; a Victrola or Edison; two or more papers; and number of languages spoken in the home. (This scale has been reconstructed by Dr. Sims[5]).

4. Composite grade in psychology. The grades in psychology were based entirely on the results of six objective tests covering all aspects of subject matter covered and set in a large variety of forms.

C. *The Nature of the Course.*

The nature of the course is described in full in the introduction of the book[6] which is the outcome of consciously planned efforts to psychologize the instruction of the first course in psychology. The lessons are intended to serve as a guide to the intelligent use of many books as source materials for active learning. The objective tests with the norms included in these lessons were used to motivate learning and an alternate series of objective tests was used for measurement.

[4] Chapman, V. C. and Sims, V. M. *The Quantitative Measurement of Certain Aspects of Socio-economic Status.* Journal of Educational Psychology, September 1, 1925, Vol. XVI, No. 6, pp. 380–390.

[5] Sims, V. M., *The Measurement of Socio-Economic Status.* Public School Publishing Co., Bloomington, Illinois, 1928.

[6] *Developed Lessons in Psychology, Including Objective Tests with Norms,* by H. Meltzer and E. M. Bailor. Harcourt, Brace and Company, 1929.

The discussion group method of teaching was used after the first week of the course. The size of discussion groups ranged from 6 to 10; leaders were changed weekly.

II. EXTENT OF KNOWLEDGE OF PSYCHOLOGY POSSESSED BY STUDENTS BEFORE AND AFTER THE FIRST COURSE

The extent of knowledge the students possessed before and after the first course was determined by an analysis of the scores made by them on the same "final" examination taken before and after the course. The highest possible score was 100. Measures of central tendency and variability "Before" and "After" were:

	MEAN	S.D.	HIGHEST SCORE	LOWEST SCORE	RANGE
Before	24.48	11.12	61	1	60
After	67.12	12.83	89	33	56

No scores on "Before" equal or exceed the mean score on "After." All scores on "After" equal or exceed the mean score on "Before." Only one score on "Before" equals or exceeds the 25th percentile on "After."

These results differ widely from those of Worcester[7] who reports that at the beginning 4 students of 185 or 2.2 per cent. receive a score equal to or in excess of the score which would be treated as a passing mark on the same test used as a final examination; that four per cent. of the class made a score at the beginning in excess of the mean for the class at the end; and that two scores exceeded the 85th percentile of final ones. On the "Before" test only one person made a score that was barely passing. Three received a grade at the beginning equivalent to an "E" or condition if the norms on the "After" examination were used for grade determination. Since Worcester does not describe the personnel of the class

[7] Educational Administration and Supervision, February, 1926, Vol. XII, No. 2, pp. 73–79.

it is difficult if not impossible to explain the differences. The most plausible explanation in the opinion of the writers is to be found in the different composition of the two groups of students. Worcester's group was in a teacher-training institution whereas of the present group less than 20 per cent were majoring in education—the remaining students were registered in Home Economics, Electrical Engineering, Agriculture, Pharmacy, and Music. It seems safe to assume on the whole, in spite of comments to the contrary, that students who major in education or enter teacher-training institutions are substantially more interested in human behavior, have read more about human behavior, and are more likely to have had courses wherein the subject matter overlaps the content of the psychology course as given in a teacher-training institution. If this explanation is valid then Worcester's conclusions about the large measure of tautological teaching of educational psychology in teacher-training institutions, but which do not hold for our group, are not necessarily applicable for general psychology courses in institutions other than teacher-training ones. To what extent they are applicable to schools of education in a college or university we cannot say. In this experiment the mean score of students majoring in education is 29.3. This is roughly five points higher than the average for the group as a whole.

III. NATURE AND EXTENT OF RIGHT AND INCORRECT RESPONSES BEFORE AND AFTER THE COURSE

The subject matter of the course was presented under five large topics. An analysis of the right, wrong, and omitted responses classified under these divisions will throw some light on the nature as well as the extent of the subject's knowledge of psychology. The content of these divisions can to a large extent be inferred from the titles.

The distribution in terms of per cent. of the right responses on "Before" and "After" examinations are classified under the five topics in table 1.

TABLE 1
Percentage distribution of correct responses before and after the course

TOPIC	NUMBER OF ELEMENTS	BEFORE	AFTER	DIFFERENCE (A–B)
Scope and Methods	17	45.56	85.40	+39.84
Study of Reactions	18	49.50	75.50	+26.00
Innate Equipment	30	48.80	70.10	+21.30
Emotions	17	40.25	75.30	+35.05
Sensations and Feelings	18	42.42	65.05	+22.63
Total	100	45.30	74.27	+28.96

1. 45.30 per cent. is the average number of right responses on the test taken as a whole when given "Before" and the variability of this average over the various aspects ranges from 40.25 per cent. in the topic *Emotions* to 49.50 per cent. in *Study of Reactions.*

2. 74.27 per cent. is the average number of right responses on the test taken as a whole given "After" and the variability is much larger than it was when given "Before"—ranging from 40.25 per cent. in the topic *Emotions* to 49.50 per cent. in *Study of Reactions.*

3. The average per cent. of increase in number of right responses from "Before" to "After" is 28.96. The largest increase is in *Scope and Method;* the next largest is in *Emotions;* and the smallest is in *Innate Equipment.*

A more detailed view of the nature of some of the right responses can be obtained by an analysis of the specific elements which were most often answered correctly on the examination taken at the beginning of the course. For that purpose let us study the questions which were correctly answered by at least 75 per cent. of the individuals. On the "Before" test there were 20 such elements out of a total of 100; on the "After" test there were 69 such elements. Of the 20, 7 are classified under *Scope and Methods;* 4 under *Reactions;* 4 under *Innate Equipment;* 5 under *Emotions;* and only 1 under *Sensations and Feelings.* The single element which

was answered correctly by the largest number of persons was a true-false question from the lesson on emotions which read as follows: "Under emotional stress man frequently displays extraordinary strength and endurance." None of the students who answered this correctly knew anything at all about the connection of this statement with Cannon's idea of emergency emotions, the neurological basis, or its social implications. They knew it as a fact of individual experience. The other questions which were answered correctly by more than 75 per cent. of the individuals were largely of this nature;—i.e., knowledge "picked up" from living in a world of human beings who are understood rather vaguely. Characteristic of the twenty questions most often answered correctly are the following:

1. The man with the biggest brain is always the most intelligent man.
2. Psychology is the science of mind reading.
3. Psychology is concerned with everyday feeling, thinking, wishing.
4. Human nature is wholly determined by heredity.
5. Mind is exempt from bodily processes.
6. A thyroid dysfunction often makes for a sluggish mental activity.*
7. By original nature infants fear loud noise.

B. *Extent and Nature of Wrong Responses "Before" and "After."*

The data in table 2 indicate the extent of wrong responses, tabulated under five topics, made before and after this first course.

The largest percentage of these before taking the course (38.18) is regarding *Innate Equipment;* the smallest (20.98) is regarding *Scope and Methods.* The percentages reported under "After" are not properly comparable to those reported in "Before," since the distributions of omitted responses

* Thyroid dysfunctions are frequent and easily observable phenomena in section of country where examination was given.

TABLE 2
Per cent. of wrong responses before and after the course

TOPIC	BEFORE	AFTER	DIFFERENCE
Scope and Methods	20.98	9.00	−11.98
Reactions	31.00	22.90	− 8.10
Innate Equipment	38.18	26.26	−11.92
Emotions	22.50	19.40	− 3.10
Sensations and Feelings	29.84	31.00	+ 1.16
Total	28.50	21.71	− 6.79

"Before" and "After" show that the largest increase in knowledge is accompanied by large decreases in omitted responses.

An analysis of the elements most frequently answered incorrectly on the examination taken before the course indicates that there are eight which are answered incorrectly by at least 75 per cent. of the individuals and 17 elements which are wrongly answered by at least 50 per cent. of the group. Of the eight elements four are from *Innate Equipment;* three from *Sensation and Feelings* and one from *Reactions.* The 17 are distributed as follows: seven from *Innate Equipment;* four each from *Reactions* and *Sensations and Feelings;* and one each from *Scope and Methods* and *Emotions.* On the same examination taken after the course there was no element which was wrongly answered by 75 per cent. of the individuals and only three elements which were answered wrongly by as many as 50 per cent. of the group.

The eight elements which were wrongly answered by at least 75 per cent. of the members of the group on the examination taken "Before" have certain common properties. As a group they (a) call for specialized knowledge which is not usually obtained without study; and (b) usually call for this specialized knowledge in a relational rather than unrelated form. Concretely, this means knowing a fact or concept and such related knowledge as historical development of concepts or names of men connected with a movement or attitude.

That this is true is indicated by the fact that only one of the eight elements answered wrongly is set in true-false form and that most of them are from multiple-choice tests. Illustrative of the eight are the following four which top the list:

1. A writer who attempts to explain all human behavior in terms of instincts is (a) Freud; (b) Thorndike; (c) Gates; (d) Warren; (e) Robinson; (f) McDougall.
2. A synapse is a connection between (a) nerve and muscle; (b) two neurones; (c) sense organ and sense center; (d) neurone and sense organ; (e) association centers; (f) two axons.
3. The item least related to the others in the following list is (a) gregariousness; (b) exploration; (c) vocalization; (d) writing; (e) manipulation.
4. Experiments show that dogs and cats learn mostly by imitation.

As a rule, the 17 answered wrongly by at least 50 per cent. of the individuals have characteristics similar to those answered wrongly by at least 75 per cent. of the individuals, but have these characteristics to a less extent.

The three elements which were answered wrongly by as many as fifty per cent. of the group are: (1) Emotional responses are adaptive. (2) Experiments show that dogs and cats learn mostly by imitation. (3) Pure sensations are psychological abstractions; the results of psychological analysis.

The number right, wrong, and omitted on the first element were 15, 39, and 6 respectively on "Before," and 21, 41, and 3 on "After." On the second element the corresponding numbers were 11, 48, and 1 on "Before," and 25, 39, and 1 on "After." On the third, the numbers were 21, 21, and 18 on "Before" and 18, 30, and 17 on "After." In general, these elements do have properties in common; they are general, abstract and interpretative; a knowledge of them presupposes a knowledge of certain psychological abstractions.

C. *Responses Omitted "Before" and "After."*

Generally speaking, the individual omits elements he does not know and knows he doesn't know. Caution in taking the test is one factor influencing the extent of omitted response. Knowledge of a subject eliminates the need for taking chances. The type of test, i.e., the objective form in which question is put, influences the extent of venturesomeness. This last fact is clearly shown by the results. On the true-false form the number of omissions is small indeed. The extent of omissions on the multiple-choice form is comparatively small. The omissions on the matching form are more extensive on the "Before" test.

The distribution of omitted responses in terms of per cent on "Before" and "After" are:

TABLE 3

Comparison of responses omitted "before" and "after"

TOPIC	BEFORE	AFTER	DIFFERENCE
	%	%	
Scope and Methods	33.46	5.60	−27.86
Reactions	19.50	1.60	−17.90
Innate Equipment	13.02	3.64	− 9.38
Emotions	37.25	5.30	−31.95
Sensations and Feelings	27.74	3.95	−23.79
Total	26.19	4.02	−22.17

IV. KNOWLEDGE STUDENTS LEARN AND KNOWLEDGE THEY FAIL TO LEARN

The gain of right responses was used as a measure of knowledge students learn and fail to learn, and an analysis was made of elements which showed a gain of at least 25 from "Before" to "After." There are 49 such elements of the possible 100. To throw some light on the nature of knowledge students fail to learn an analysis was made of elements which show a gain of only 5, elements which show a gain of 10 or less, and elements which show gains of 15 or less. The

number of elements in the various divisions of psychology as presented in the course and previously described are indicated in the following table:

TABLE 4

Distribution of gains on elements students learn and fail to learn

TOPIC	GAINS OF 25 UP	GAINS OF 15 OR LESS	GAINS OF 10 OR LESS	GAINS OF 5 OR LESS
Scope and Methods	11	6	1	.....
Reactions	5	7	4	4
Innate Equipment	9	8	4	.....
Emotions	11	5	4	.....
Sensation and Feelings	13	2	2	3
Total	49	28	15	7

SOME GAINS DUE TO CORRECTION OF SOME MISCONCEPTIONS

In his recent novel, *The Age of Reason,*[8] Gibbs has three characters express three different attitudes toward beliefs: 1. Perhaps a little superstition is better than unbelief (Halliday, page 146); 2. Any decent superstition rather than blank unbelief (Cyril, page 181); 3. There are so many things to believe without being superstitious (Cannon Allport, page 196).

An analysis of the nature of the specific elements students learn readily as shown by large gains indicates that by and large the students in the course favor the last expressed attitude. Some of the largest gains are the result of a re-learning process, including the unlearning of such misconceptions or misstatements as "Fear of snakes is native," "Man instinctively reaches for the telephone to report a fire," or "Anger as a response to insulting language is native," followed by learning the facts as psychology teaches them.

MOST GAINS DUE TO ACQUISITION OF SPECILIZED KNOWLEDGE

Most large gains are, however, due to the acquisition of specialized knowledge; and the elements which show most

[8] Gibbs, Philip—*The Age of Reason;* Doubleday, Doran, and Co.

gains have characteristics similar to those described in the analysis of knowledge not possessed, *i.e.*, specialized knowledge not obtained merely by living long enough.

An analysis of the number of right responses on "Before" and "After" of the elements which show small gains indicates that the smallness of gain in approximately one-third of the cases is directly explicable by the fact that most students knew them before they took the course and hence large gains were impossible. Those elements already listed as answered wrongly on the test taken after the course are characteristic of those which students did not answer correctly on the test taken "Before" and also failed to learn. The following also are such elements:

1. Some native traits are not universal.
2. The more complex a reaction is the more it is susceptible to central modification.
3. Acquired traits are dependent on native traits.

The general characteristics of these questions are much like those described under knowledge not possessed after the course. They are general, interpretative, and call for relational thinking.

V. RELATIONSHIP OF GAIN IN KNOWLEDGE OF PSYCHOLOGY TO INTELLIGENCE, SOCIO-ECONOMIC STATUS, AGE, AND GRADE IN PSYCHOLOGY

In the calculation of the relationship between gain in knowledge and other measurable factors, raw gain in the number of points made on the 100 element examination taken before and after the course was taken as a measure of gain in knowledge. Intelligence and socio-economic status were measured as previously described. By grade in psychology is meant the total score on six objective tests given throughout the quarter. The correlations were calculated by the Pearson Products-Moment formula:

Gain in psychology	*"r"*	*P.E.r*
1. Intelligence	+.099	±.14
2. Socio-economic status	−.017	±.15
3. Chronological Age	−.210	±.11
4. Grade in Psychology	+.420	±.078

These coefficients are lower than those which usually are obtained. The correlation of $+.42 \pm .078$ with the objective achievement tests indicates that there is a tendency for those who make the highest scores on a series of tests distributed over the quarter to make slightly greater gains. In a sense it may be taken as an indication of the extent of consistency on the part of the individuals in the group to score on par with their usual combination of ability and effort.

VI. SUMMARY AND CONCLUSIONS

This article reports a study made of the extent and nature of knowledge of the psychology that students possess or do not possess, learn or fail to learn before and after the first course. As a measure of knowledge of psychology a "final" examination composed of 100 elements and set in three different forms was given to a group of students on their first appearance in class and again on their last appearance. Mental, educational and social data were obtained by the use of tests, measures, and questionnaires. The results and some of the conclusions drawn from them may be summarized as follows:

1. An analysis of the distribution of scores made by students before and after the course yields results which indicate the extent of their knowledge. The arithmetic mean of the test taken "Before" is 24.48 and on the test taken "After" 67.12. Further analysis reveals results which differ widely from those found by Worcester. From the data presented it is concluded that Worcester's conclusions concerning the vast amount of tautological teaching in higher education are not applicable to our group and by extension, possibly, not wholly applicable to courses in general psychology in other than teacher-training institutions.

2. The nature of the knowledge of students was studied by analyzing the extent and nature of their right, wrong and omitted responses before and after the course. Such an analysis shows among other things:

a. In terms of divisions of subject matter as described in the article, the largest gains are made in *Scope and Method* and *Emotions,* whereas the smallest are made in *Innate Equipment* and *Sensations and Feelings.*

b. Respecting the knowledge of psychology on the part of students before taking the course, it appears that some knowledge is the result of individual experiences, *i.e.,* knowledge "picked up" from living in a world of human nature which is understood rather vaguely and with little certainty. The uncertainty is indicated by the fact that a change in the form of the test from true-false to any other, results in more omissions.

c. An analysis of wrong responses before and after the course indicates:

(1) The largest per cent of misconceptions *before* taking the course is from *Innate Equipment;* the largest *after* the course from *Sensations and Feelings.*

(2) Questions answered incorrectly most frequently call for specialized knowledge which is not easily obtainable merely by living and require that knowledge to be relational rather than discrete and unrelated.

(3) The elements which students fail to learn have common characteristics: they are general, abstract, interpretive, are not likely to be set in true-false form, and a knowledge of them presupposes ability to grasp the meaning of such abstract concepts as "pure sensations," "psychological abstractions," "adaptive," and "imitation."

d. An analysis of test elements omitted before and after the course indicates that:

(1) The smallest per cents of questions omitted before taking the course are from *Innate Equipment* and the largest

are from *Emotions*. The largest and smallest differences in per cent. omissions from "Before" to "After" are in reverse order.

(2) Venturesomeness, or willingness to take chances, affects the extent of omission; knowledge possessed and the form of test influence venturesomeness.

3. The extent and nature of knowledge students learn and knowledge they fail to learn were studied by analyzing the gains made from "Before" to "After." Such an analysis shows among other things that:

a. The largest gains are in test elements from *Sensations and Feelings;* the smallest from *Reactions*.

b. A large number of gains are due to the unlearning of superstitions, misconceptions, or misstatements made before taking the course.

c. Most gains are due to the acquisition of specialized knowledge.

4. A study of the relationship of the gain in knowledge in psychology with certain factors was made and the following results obtained:

a. Within the limits of the present data gain in knowledge of psychology is not conditioned by intelligence. This is indicated by a correlation of $.099 \pm .14$ between these variables.

b. Socio-economic status correlates $.17 \pm .15$ with gain in knowledge of psychology from "Before" to "After."

c. Gain in knowledge of psychology correlates $-.21 \pm .11$ with chronological age. Were it not for the narrow range of age variability the correlation probably would have been more significantly negative.

d. With the six objective tests in psychology given at distributed periods during the session, gain in knowledge of psychology correlates $.42 \pm .078$. This may be taken as somewhat of a measure of the consistency of the individuals in the group to score on par with their usual combination of ability and effort.

A PRELIMINARY STUDY OF THE AFFECTIVE VALUES OF COLORED LIGHTS

WILLIAM E. WALTON
University of Nebraska
AND
BEULAH M. MORRISON
University of Kansas

Introduction. Although there is a mass of material relative to the affective values of pigment colors of various sorts, there is no published work on the affective values of colored lights so far as the writers have been able to determine. The experiments reported in this article were designed for the following purposes: (1) to determine whether or not students at the University of Kansas have preferences for colored lights; (2) to establish scale values for these preferences; (3) to investigate the constancy of the choices; (4) to discover any possible sex differences; (5) to learn what factors are responsible for these selections.

Procedures. Preferences for certain colored lights were obtained by the method of paired comparisons from about three hundred forty students in elementary psychology. Lights from five Mazda "natural" colored bulbs, *i.e.*, those in which the glass itself is colored red, green, blue, amber (yellow), and clear (white), were presented to the students by means of the chromopathometer,[1] an apparatus designed and constructed by one of the writers (W), which throws discs of diffused light upon a frosted glass screen. Single colors and combinations of two colors were presented and judged by the method of paired comparisons. Although the experiments were performed in a semi-dark room, sufficient

[1] For description of this apparatus see *J. Exper. Psychol.*, 1929, 12, 254–258.

light was admitted to enable the students to mark their choices on systematically arranged record sheets from which the results could be summarized easily.

There were four different experiments involving two groups of students, one group of about 165, the other of about 178. The number of students for single colors and for combinations differed slightly. The first experiment was performed April 13, 1927, and was repeated (designated as second experiment) with a part of the same observers May 16, 1927, in order to secure data on the constancy of the choices. Another group was asked to make similar choices (third experiment) October 11, 1927. On January 17, 1928, a large number of these observers reported preferences when the intensities of the lights were equated (fourth experiment). The relative brightness of the original lights from the least to the greatest is indicated by the following order: blue, red, green, amber, clear. With the intensities equated, the clear and amber lights were very much alike because of the yellow tinge in the so-called clear light.

Immediately preceding the third experiment, the observers were asked to state in answer to specific questions what colors and combinations of colors they especially liked and disliked and why.

Treatment of Data. The data were separated according to sex only, since the observers were a fairly homogeneous group with respect to age, intelligence, race, social status and the like. First, a rough summary of the data was secured by totaling the preferences for each color. Then to get a relative evaluation of the choices, the method developed by Guilford[2] was applied. This method gives the scale values in terms of *sigma* units from the mean of all the stimulus values. The average per cent of times each color was preferred in all its comparisons was found by the formula,

[2] J. P. Guilford, The method of paired comparisons as a psychometric method, *Psychol. Rev.*, 1928, 35, 494–506.

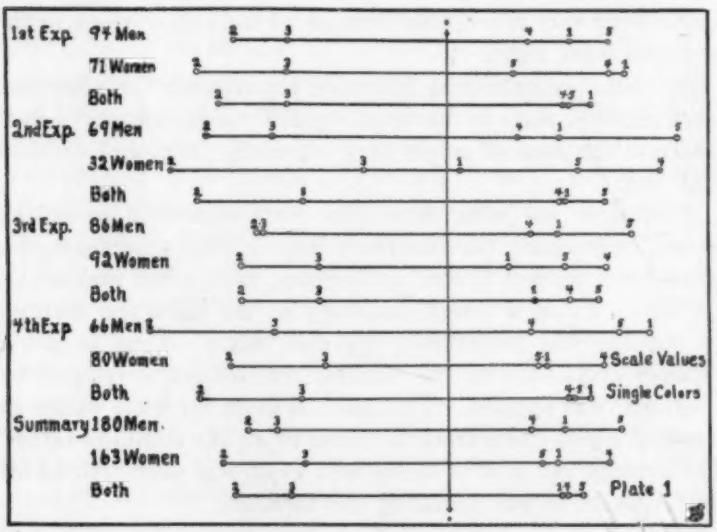
$$\text{Av. Pc.} = \frac{N_k + N_s/2}{N_s N_r}$$

where N_k is the total number of choices given to a color,

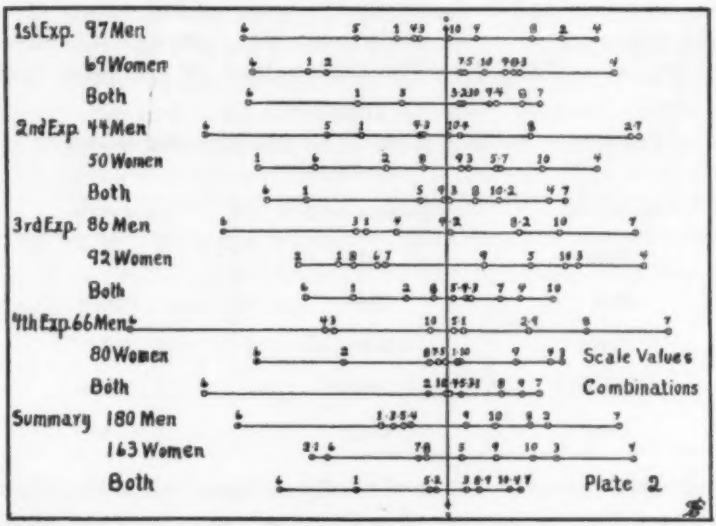
N_s is the number of observers,

N_r is the number of colors.

The Av. Pc. minus .50 then gives the distance of the stimulus color from the standard of the scale in terms of per cent. The scale value was found by reference to a fundamental table giving deviates and ordinates of the normal probability curve in terms of area from the mean, such as is found in Holzinger's tables.[3] For purposes of direct comparison throughout the treatment of these data, the scale values are plotted along lines from the same zero point, the larger values to the right and the smaller to the left (Plates 1 and 2). The numbers, one through five, along the scale lines refer respectively to red, clear, amber, green, and blue; along the scale lines for combinations the numbers, one through ten, refer similarly to



[3] K. J. Holzinger, *Statistical Tables for Students in Psychology and Education*, 1925, Table XII.



the combinations red-amber, green-blue, amber-green, clear-green, red-clear, clear-amber, red-blue, red-green, amber-blue, and clear-blue.

Results. Throughout the investigation the saturated single colors were preferred. This is in agreement with the findings of many investigators who have worked with *pigment* colors, among them Jastrow,[4] Luckiesh,[5] Bradford,[6] but in disagreement with the findings of others such as Major,[7] and Washburn.[8] The combined results of the first and third experi-

[4] J. Jastrow, The popular aesthetics of color, *Pop. Sci. Mo.*, 1897, 50, 361–368.

[5] M. Luckiesh, A note on color preference, *Amer. J. Psychol.*, 1916, 27, 251–256.

[6] E. J. G. Bradford, A note on the relation and aesthetic value of the perceptive types in color appreciation, *Amer. J. Psychol.*, 1913, 24, 545, 554.

[7] D. R. Major, On the affective tone of simple sense impressions, *Amer. J. Psychol.*, 1898, 7, 57–77.

[8] M. F. Washburn, A note on the affective values of colors, *Amer. J. Psychol.*, 1911, 22, 114–115.

ments involving 343 observers, 180 men and 163 women, give the following order of preference: blue, green, red, amber, clear. (See Table 1.) The scale values for the three most

TABLE 1

Summary of the rank orders of the affective value of single colored lights

MEN—180	WOMEN—163	COMBINED
Blue	Green	Blue
Red	Red	Green
Green	Blue	Red
Amber	Amber	Amber
Clear	Clear	Clear

preferred, however, are very nearly the same, while there is a great difference between this group of colors and the other two colors (Plate 1). Comparing this order with the combined results of the experiment with the intensities equated, we find that red shifts to first place, which would indicate that under the conditions of this experiment lowering its intensity makes red more pleasing (Plate 1).

There is some difference between the preferences of men and of women for the three saturated colors, the order for men being blue, red, green, amber, clear; for women, green, red, blue, amber, clear. (See Table 1 and Plate 1). These preferences of the men remained the same throughout the investigation except when the intensities were equated in the fourth experiment, in which case red was found to be most preferred. The choices of the women were not so constant. In the first experiment red was most preferred; in the second (repetition with a small number, 32, of the same observers) red dropped to third place, while green took first, an order which was maintained in the third experiment; in the fourth (intensities equated) red shifts to second place. These changes in the order of preference may be traced by the numbers on the scales of Plate 1.

It would be well to know the statistical reliability of each scale value in relation to every other scale value, but in a preliminary study the number of calculations necessary would not justify the end to be attained. However, we have measured to a degree the reliability of the positions of red and blue by computing the probability of a positive difference between them, both under the normal intensities and under equated intensities. This also tells us the probability of a shift in the order of red and blue under the same conditions. The following formulae were used:

1. Diff., in P.E. units equals $\frac{\text{Diff. of P.C.}}{\text{P.E. of Diff.}}$
2. P.E. of Diff. equals $\sqrt{\text{P.E.}^2_{\text{p.c.}_1} \text{ plus } \text{P.E.}^2_{\text{p.c.}_2}}$
3. $\text{P.E.}_{\text{p.c.}}$ equals $.6745\sqrt{\frac{pq}{n}}$

The accompanying table presents the results of these calcu-

TABLE 2

Reliability of differences in scale values of red and blue under normal and equated intensities for men and women

SEX	COLOR	POSITION IN SCALE	COLOR	POSITION IN SCALE	NORMAL OR EQUATED INTENSITY	CHANCE THAT CHANGE IN ORDER WILL OCCUR
Women	Red	3	Blue	2	N. I.	1 to 22.2
Women	Red	2	Blue	3	E. I.	1 to 1.4
Men	Red	2	Blue	1	N. I.	1 to 65.0
Men	Red	1	Blue	2	E. I.	1 to 35.0

lations. It would seem that the scale values for red and blue as determined for men are reliable as far as the conditions of this experiment are concerned, but for women the reliability is low. This also means that the shift in order of preference for red and blue when the intensities are equated is significant in the case of the men, but is of doubtful significance in the case of the women.

A further test of reliability of the scale values is found in the fact that of the 165 observers of the first experiment with single colors, 77 per cent (75 per cent of the men and 79 per cent of the women) chose one particular color every time that it appeared.

The combinations most preferred and those least preferred stand out rather clearly. There is, however, a great deal of shifting in the middle ground, and men and women differ considerably in their preferences. In general, men prefer combinations of saturated colors; red-blue is always first, while green-blue and red-green appear frequently in second and third places. (See Table 3 and Plate 2). On the other hand,

TABLE 3

Summary of the rank orders of the affective value of combinations of colored lights

MEN–183	WOMEN–161	COMBINED
Red–Blue	Clear–Green	Red–Blue
Green–Blue	Amber–Green	Clear–Blue
Red–Green	Clear–Blue	Clear–Green
Clear–Blue	Amber–Blue	Red–Green
Amber–Blue	Clear–Red	Amber–Blue
Clear–Green	Red–Green	Amber–Green
Red–Clear	Red–Blue	Green–Blue
Amber–Green	Clear–Amber	Red–Clear
Red–Amber	Red–Clear	Red–Amber
Clear–Amber	Green–Blue	Clear–Amber

women prefer a combination of one saturated color and one unsaturated color; clear-green is first except when the intensities are equated, in which case amber-green moves to first place, probably because of the similarity of amber and clear under those conditions. The combinations, amber-green and clear-blue, are found most frequently in second and third places. There are, however, two experiments in which a combination of two saturated colors appears in third place. These exceptions move to sixth and seventh places in the summary of results (Table 3).

The least preferred combination is clear-amber. This is not surprising, for these lights are so much alike. Men choose the saturated-unsaturated combination least often, while women tend to dislike the two-saturated color combination.

The results for men seem to bear out Geissler's[9] proposed rule that the greater the pleasantness of the constituents, the greater will be the pleasantness of the combination; their preferences for single colors being in order blue, red, green, amber, clear (intensities normal); the three most preferred combinations being red-blue, green-blue, and red-green. But for women this is not the case; their single preferences are in order green, red, blue, amber, clear, but for combinations the first four preferences are clear-green, amber-green, clear-blue, and amber-blue (Plates 1 and 2). Obviously in this group there is a sex difference, and undoubtedly other factors than a summation of constituent pleasantness play a part.

Reliabilities were not calculated for any combinations, but by following the combinations through Plates 1 and 2 one may gain a general idea of the degree of reliability of these choices. An analysis of the preferences of the first experimental group shows that 48 per cent of the whole group (48.5 per cent of the women and 47.8 per cent of the men) chose a particular combination every time it appeared.

A random sample of fifteen sets of answers to the questions asked preceding the third experiment was analyzed in order to obtain an indication of the relationship between the colors the observers said they preferred and the colors actually chosen first in the experiment. Eleven of the *O*'s, or 73 per cent, gave first place in the experiment to the single color of the same name as the one they said they preferred, while six *O*'s, or 40 per cent, preferred the same combination in both instances.

A complete analysis of all the answers given by the 178 observers throws considerable light on the factors involved in

[9] L. R. Geissler, The affective tone of color combinations, *Studies in Psychology*, Contributed by Colleagues and Former Students of E. B. Titchener, 1917.

these choices. The reasons for liking and disliking colors or combinations may be classified roughly as follows:

1. They suggest clothing worn by the observer or his acquaintances.
2. They suggest or arouse an emotional state, *e.g.*, soft and restful or dismal and lifeless.
3. The harmony or contrast of the two colors is pleasant or unpleasant.
4. They suggest out-of-doors or seasons.
5. They are over-popular.
6. They resemble colors of organizations such as classes, clubs, schools.
7. They recall specific experiences or have symbolic meaning, *e.g.*, accident, blood, cleanliness.
8. The choice is due to personal taste or feeling, conditioning factors have been forgotten.

Undoubtedly social conditioning including conventional and cultural associations, and individual experiences are factors of great importance in color preference. Perhaps they are the main factors. Many investigators, among them Lillie,[10] Mercer,[11] Pressey,[12] Slade,[13] and Washburn,[14] have pointed out the social factor. The gross agreement within groups and races seems to be largely the effect of common social influences, while the ever present individual differences are conditioned by individual experience. An individual's preference may change according to his personal experiences, the vogue of

[10] W. Lillie, A note on the nature of color associations, *Mind*, 1926, 35, 533–536.

[11] E. M. Mercer, Color preference of one-thousand-six negroes, *J. Comp. Psychol.*, 1925, 5, 109–146.

[12] S. L. Pressey, The influence of color on mental and motor efficiency, *Amer. J. Psychol.*, 1921, 32, 326–356.

[13] T. K. Slade, An inquiry into the nature of color association, *Mind*, 1925, 34, 455.

[14] M. F. Washburn, and D. Crawford, Fluctuations in the affective value of colors during fixation for one minute, *Amer. J. Psychol.*, 1911, 22, 579–582.

the season or decade, his knowledge of theories of color harmony and the like. The present day internationalization of culture will tend to make color preferences alike the world over. The American becomes accustomed to the colors and color combinations of the western European peasant, the Slav, the Chinese, the East Indian, and applies them in his art and industry, while the European and the Oriental modify their preferences and practices in accordance with western ideas.

Summary. This study (1) presents scale values for preferences for five colored lights and their combinations among 340 students at the University of Kansas. (2) The preferences of the men and of the women differed somewhat both for single colors and for combinations. (3) Within the scope of these experiments the choices remain only fairly constant, those of the men being somewhat more constant than those of the women. (4) It is quite clear that a multiplicity of factors is responsible for these preferences, since equating the intensities of the lights affected the choices, and an analysis of the reasons for color preferences given by the students indicates that many social conditionings are significant.

THE NECESSITY OF PSYCHOMETRIC TESTS IN THE STUDY OF MALADJUSTMENTS[1]

HARRY W. CRANE
University of North Carolina

The thesis I wish to present here is that psychometric findings with standardized tests *together with* the qualitative results as indicated both by the character of the responses and the general method of reaction to the various elements of the test are in many cases absolutely necessary to an adequate understanding of maladjustments. The qualitative responses and the methods of reaction to the test situations are the replicas of the method of reaction to the ordinary life situations of the patient.

The following case is partially cited in summary to illustrate the above points:

Patient: Male, age 21. Junior in college, department of commerce. College work passing but patient uninterested.

Family: Father dead. Mother unusually intelligent and active. Cultural background; social and financial status, good. Evidences of neural instability in alcoholic and drug excesses of two maternal aunts.

Maladjustment Symptomatology: The passing of worthless checks; gambling with group of social inferiors to whom he always lost. Mild mental depression; melancholy; unresponsive; careless about personal appearance and rights of others; extremely uncommunicative with every one. Manner brusk and repellent; no intimate friends.

History: The only suggestive data secured from a case history was that the patient had been unjustly reprimanded and expelled from a preparatory military school for stacking of

[1] This paper was given in part at the Symposium on Maladjustments at the Ninth International Congress of Psychology.

the dormitory of which he was not guilty. He had been told that he would not be punished if he confessed and apologized. He preferred punishment to false confession. The mother states that the patient seemed extremely bitter and disheartened over this treatment. She thinks she noticed following this incident the beginning of the character change from a normal, happy, communicative boy to the sullen, uncommunicative type now pictured.

Patient's General Reactions During Psychological Conferences: Opposed to the apparent seclusiveness indicated by the history the patient seemed perfectly willing to answer all direct questions asked by the consultant, even those relating to sexual irregularities. However it was impossible to get any straightforward, free-worded responses to any general questions or detailed narrative accounts of experiences.

Without data from formal psychological tests the following tentative statement about the case would have been: the maladjustment difficulties are the result of unfair treatment from school authorities, from this treatment originated the general antagonistic attitude toward authority and society with the consequent flaunting of disapproved behavior in the face of social codes. Together with this cause is the failure to see the value of correct behavior, since one suffers anyway. The consequences are the gambling, unreliability in money matters, bruskness of manner and seclusiveness. The apparently frank answers to direct questions on consultation are to be interpreted as masking the real seclusiveness, evidenced by the inability of the examiner to get free narrative statements.

Indications from Test Data: Among other interpretative data yielded by tests was a mass of evidence bringing out the presence and importance of a marked difficulty in comprehending and use of language. The following is a brief summary of the more important evidence.

On the Stanford Revision of the Binet-Simon: Basal Age, 8; Mental Age, 15/11; Intelligence Quotient, 100. The most striking of all the data was a vocabulary score of 57 (14-year

level). Further indicating inferior language ability was having the only failure at the 14-year level to occur in the test of the group most dependent upon language ability, namely, problems of fact; coupled with this are a number of successes at this and higher levels with difficult tests less dependent upon the language factor—arithmetical reasoning, reverse clock hands (time for the three parts being 13″, 11″, and 16″), code test (one error, time 2′ 51″), paper cutting and the 18-year-old level ingenuity test.

Stanford Achievement Test: On the six subjects covered in this test, the highest rating of the subject was on arithmetic with a Subject Age of 18. This is the test in the group least dependent upon language. Practically no words are used in the comprehension part of the test, while the section on arithmetical reasoning is given in very simply formulated sentences. The Subject Ages for the other sections of the test were: Reading, 16/11; History and Literature, 16/4; Language Usage, 16/8; Nature and Science, 16/8, and finally Dictation Spelling, 14/7. From further analysis of the reading results we find that the lowest Subject Age, 16/3, was obtained in that part of the test made up of the more intricate language mechanisms, namely, paragraph meaning. While the Subject Ages for the simpler reading mechanisms, sentences and word meaning, were 18/6 and 17/2 respectively.

Roback Mentality Test for Superior Adults: The particularly significant points in this test are that with the exception of the Cryptogram Test which is very little dependent upon language and in which the patient scored 100 per cent, all the other tests are decidedly dependent upon language ability. The one of the series (Acumen Test) most dependent upon this type of ability is the one in which the patient scored only 6.7 per cent. The instructions for this test are: "State in very concise and convincing language what is logically wrong with each of the following statements and state why." Example: "If I am not mistaken, I met them last week."[2]

[2] Below are given the percentage of possible scores made on each of the sub-tests of the Roback Test:

Sub-Test	%	Sub-Test	%
Cryptogram	*100.*	Abstraction	41.6
Directions	71.4	Subsumption	41.3
Relations	66.	Insertion	33.3
Analogy	58.	Problem	32.
Reference	43.	Judgment	30.
Opposites	41.7	*Acumen*	*6.7*

Stenquist Mechanical Aptitude, 1 and 2; and Gwinn-Thurstone Ingenuity: Both of the Stenquist tests involved the minimum use of language in giving and none on the part of the patient. His percentile ranks were 84 and 89 in these two tests. Contrasted with these ratings, a percentile rank of 21 was obtained on the Ingenuity test where nearly all of the problems are stated in relatively complex language.

On the basis of these findings we concluded that the patient's apparent shut-in-ness was not at all of the introvert type. Also, that while it might be partly determined by his early unfair treatment, it was at least intensified and perpetuated by the difficulty in communication. As additional evidence against introversion was an ability and decided interest in music together with a participation in a college and in a commercial band. Here we see the possibility of the patient's adjustments in forms of social cooperative enterprises of the type in which the language factor is of minor importance. The bruskness of manner and the lack of friendships, the seeking of a group of companions of lower status with whom he gambled are probably to be considered as defense mechanisms for the purpose of hiding his language difficulty. That the subject sensed this difficulty became evident upon direct questioning.

The data from the Downey Will-Temperament Test (Fig. 1) give findings of a confirmatory type. Against the impression of extreme antagonism given by the patient's uncommunicative, brusk and repellent manner, we have the Downey score of 6 for Resistance to Opposition and of 7 for Reaction to Contradiction. Neither of these scores is markedly above the average.

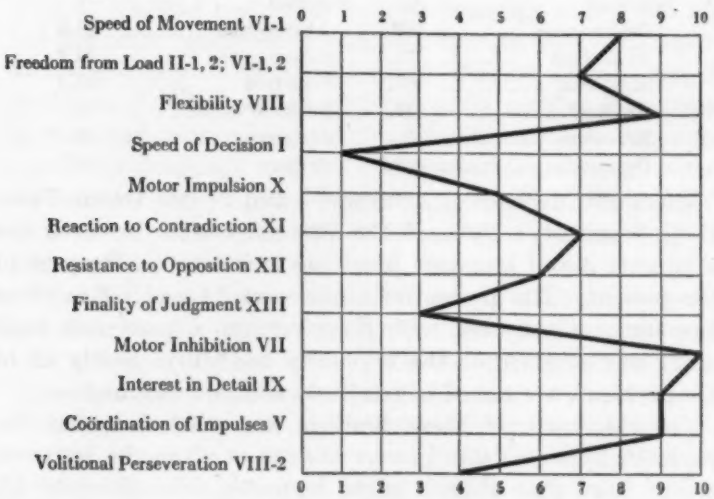

Speed of Movement VI-1
Freedom from Load II-1, 2; VI-1, 2
Flexibility VIII
Speed of Decision I
Motor Impulsion X
Reaction to Contradiction XI
Resistance to Opposition XII
Finality of Judgment XIII
Motor Inhibition VII
Interest in Detail IX
Coördination of Impulses V
Volitional Perseveration VIII-2
0 1 2 3 4 5 6 7 8 9 10
0 1 2 3 4 5 6 7 8 9 10

FIG. 1. DOWNEY WILL-TEMPERAMENT PROFILE.

The patient showed in many of his test reactions evidences of his having formed a general habit of careless, inconsiderate reaction to those situations not particularly interesting to him. Note especially the high speed and low accuracy on the Benge Clerical-K (77 per cent speed; 36 per cent accuracy). This is also borne out by the fact that he finished the Roback tests quickly but with lack of thoroughness. In line with this behavior are the Downey scores of 8 on Speed of Movement and 9 on Freedom from Load.

The extent of the patient's self-consciousness and shyness together with the important part they contributed to his maladjustment are made quite evident in the Pressey Cross Out Test. In the Unpleasant word section of this test there were checked 72 per cent of the disgust, fear and self-feeling words. These are in reality all self-feeling words. Only 48 per cent of the sex words were checked. Again in the Worry section the predominance of the checking for selfconscious words, 44 per cent, is significant. The percentages for the other types

of words are: Hypochondriacal, 20 per cent; melancholy, 36 per cent; neurotic, 28 per cent; paranoid, 8.4 per cent.

The test situations are seen to offer excellent means of determining the presence of a general habit pattern duplicating the indication of the same kind of behavior in the patient's regular life situations; namely, his carelessness about clothes, money matters, and responsibilities.

It is easy to see how inadequate would have been the interpretation of the maladjustment symptoms without the additional formal test data. The material presented here also brings out that the giving of numerous tests makes possible the eliciting through different formulations and materials of characteristic defects in methods of meeting life situations as well as in capacities.

THE AESTHETICS OF RECTANGLES AND THEORIES OF AFFECTION

C. O. WEBER
Wells College

In a previous article[1] an account was given of Moore's significant studies of consonance and dissonance and of my own studies of visual forms. Both of these studies are in harmony with the so-called "conative" theory of affection. In the present paper I will report another study confined to the feeling-tone of rectangles. In addition to the golden section, we employ as material four new luminaries in the field of art known as the "root rectangles." The essential part of our procedure that is unique consists in presenting the materials a second time for the purpose of testing the presence of a certain shift in the judgments which we were led to expect from the results obtained by H. T. Moore in the sphere of tone.[2] The results of the present experiment satisfy the expectation, and serve as experimental support for the theory of affectivity which Moore develops in connection with his findings.

Nine rectangles were used, consisting of the golden section, four root rectangles, and four interpolated rectangles, whose proportions are such that they serve as transition forms between the root rectangles. If these rectangles are arranged according to the ratio of width to length, they vary from a square (No. 7) to the root five rectangle (No. 23) which is narrow and long. The area of each rectangle is 20 square centimeters. The ratio of the golden section is familiar

[1] Weber, C. O., "Theories of Affection and Aesthetics of Visual Form," *Psych. Rev.*, Vol. 34, 1927, pp. 206–219.

[2] Moore, T. H., "The Genetic Aspect of Consonance and Dissonance," *Psychol. Monog.*, 17, 1914, No. 73.

(width is to length as length is to sum of width and length); but the root rectangles possess a more indirect type of symmetry. With the root rectangles, not the width and length as such, but their squares, are in ratio. Thus, the square erected on the width of the root two rectangle is to the square of its length, as one is to two. Similarly, the base and height of the root five rectangle, after squaring, are in the ratio of one to five. If it is true that these rectangles are favored as aesthetic forms, it implies that the observer is influenced, not by the sides themselves but the outlying areas which they seem to delimit. That the eye should consider bounding areas in judging a given figure does not strain credulity; but it does tax belief to be told that the eye is skillful enough to determine without aid that the squares of two sides of a rectangle are in simple ratio! Our results, in fact, indicate that the observer exhibits no such surprising capacity. Of course, average observers may not be aesthetic observers; but we are concerned here not with beauty but with affective-tone. Students of aesthetics claim that the root rectangles have been favored in art, that they are favored proportions found in the painting, sculpture and pottery of famous arts, beginning with those of Greece.[3] Modern students give added significance to the root rectangles by offering proofs that their ratios occur naturally in the formation and distribution of leaves, and in other spheres of nature.

The following list gives, after the numbers of the rectangles used, their names and the ratios of width to length. Only odd numbers are used as means of identification to avoid the tendency to favor even numbers when judgment is in doubt.[4]

[3] For an excellent and beautifully illustrated account of the root rectangles in the history of early arts see Jay Hambridge's *Dynamic Symmetry*, 1920.

[4] Such tendencies to favor certain numbers would of course be slight. But Coover's well-known experiments in psychic research and C. O. Mathew's study of the effects of position of printed words (*J. of Ed. Psych.*, 18, 1927, pp. 445–457) seem to show that such preferences are not negligible.)

Number	*Designation*	*Ratio*
7	Square	1 : 1
9	Interpolated	1 : 1.207
11	Root two	1 : 1.414
13	Golden Section	1 : 1.619
15	Root three	1 : 1.732
17	Interpolated	1 : 1.871
19	Root four	1 : 2.000
21	Interpolated	1 : 2.122
23	Root five	1 : 2.236

These nine rectangles were presented by the method of paired comparisons to 68 college women in two experimental groups. After the lapse of two weeks they were represented under conditions as much like the first as possible. The figures were thrown on a screen always with the length in vertical position. Each student recorded the number of the rectangle preferred and a judgment of certainty or uncertainty. We give first the results of the preference judgments, postponing the question of certainty of judgments. Table 1 gives the total frequency with which each rectangle was preferred in the first and second presentations.

TABLE 1

RECTANGLE	TRIAL I	TRIAL II
7	176	134
9	223	184
11	292	259
13	340	340
15	359	378
17	365	379
19	309	349
21	222	253
23	157	180

It is clear that according to these results the golden section and the root rectangles are not a separate species or type. This is the usual fate of alleged "types" in psychology. They appear when tested in the laboratory as but tendencies or the

extremes of tendencies. Most human beings as recent studies show are neither introvert nor extravert, but combine both traits. Just so, the average observer favors neither the square nor the narrow root five rectangle but chooses an intermediate form. Even the studies of Fechner showed that the golden section only *tends* to be the modal choice. The studies of recent writers, Witmer, Angier and others, confirm this result. The golden section is perhaps not favored because of the mere mathematical symmetry of ratio it possesses but because it has a form which avoids the banal simplicity of the square, but also the too daring rectangularity of root five. In our experiment the maximum preference falls not on the golden section but on No. 17, an interpolated rectangle, without any respectability in the history of aesthetics! However, the golden section (No. 13) is an important rival.

A much more significant result is *the shift towards a more complex ratio* which occurs when the rectangles are presented a second time. This is so uniform that it can by no means be regarded as an accident. T. H. Haines and A. E. Davies[5] have emphasized the importance of considering individual as well as mass results. To this end we secured a score for each subject which represents her "tolerance" for extreme ratios. She was given one point each time she preferred No. 7, 2 points for each case where No. 9 was preferred, and so on. The results show that 55 out of the 68 subjects shift from a smaller to a large score, *i.e.*, towards the root five ratio. The average weighted score (as we may call it) for trial 1 was 179, for trial 2 it was 190, a difference of 11 points. Twelve students had lower scores for trial 2, and one student did not change at all.

Perhaps this tendency to prefer more complex ratios after practice is compensated for by a shift to more "uncertainty" in the certainty judgments. A survey shows that the total number of "uncertain" judgments drops from 822 in trial 1

[5] Haines, T. H., and Davies, A. E., *Psychol. Rev.*, Vol. 2, 1904, pp. 249–281.

to 713 in trial 2. It might still occur, however, that the complex rectangles, though they increase in preference, might lose in the certainty with which they are preferred. On the contrary the square in trial 2 was preferred with *less* certainty than in trial 1. As we move towards the root five rectangle the percent of certainty in the second trial as compared with the first *increases.* The course of the certainty judgments, therefore, strengthens the conclusion that *practice in seeing rectangles induces a demand for more daring ratios.*

What are the bearings of these results for affective theory? They are quite in harmony with my previous study (referred to), and as we shall see with the results attained by others: Moore, Pierce, and Washburn. The pleasure or agreeableness of any activity depends on the capacity of the subject. *To be agreeable, any activity in which we indulge must tax capacity but not tax it too much.* Why does preference shift to more difficult perceptions after practice? Because practice tends to increase capacity. Moore found that as subjects listen to series of chords, some consonant and others dissonant, the preference for dissonance grows with practice in listening to it. In his engaging little volume, *Pain and Pleasure* (1917), Moore extends this principle of affection to many spheres of life. The principle is a thoroughly common-sense one. That difficult physical tasks should be pleasant to the physically strong, that difficult formulas should exhilarate the mathematician, that the simple harmonies of Bach and Brahms should please the listener who is indifferent to the more complex themes of Wagner and actually repelled by the compositions of Stravinsky and Schonberg, is not surprising. In general, I think it may be said that the rôle of volition and conation is not surprising in concrete life—only psychology looks askance because it cannot do justice to volition and still ape physics.

Perception is a dynamic act: the eye does not receive impressions passively. Anything that is perceived at all must be unified and the most primitive step in unification is isolation. For example, on a dark night perception begins when

"something" is differentiated from the background. But if mere unity were able to elicit pleasure, then the square and the most harmonious chord should be the most pleasant. The square is affectively indifferent—it is too easy to apprehend. The mind wants variety added to unity, as much variety as is consistent with perceiving at all. Says Moore of musical chords, "It is the barely successful synthesis, in which the individual accomplishes with difficulty the unification of a manifold, that he finds the keenest pleasure."[6]

We turn, finally, to a brief account of some results of other experimenters which are in harmony with our own. It appears that practically no investigators have explicitly tested the tendency of practice to induce preference for more complex perceptions in the sphere of vision. In this regard an early study made by Edgar Pierce comes the nearest to resembling our own.[7] Pierce experimented with simple lines and figures. We call attention to certain results obtained with simple lines. Suppose that a black field is divided by a white vertical line. On the right is a white line 10 cm long at a fixed distance of 20 cm from the dividing line. On the left is a movable line 5 cm long. The observer shifts the short line back and forth until he obtains an arrangement of the three lines which satisfies him. The results show that observers move the short line farther away from the central line than the 20 cm which separates the central from the other line. Pierce says that it is as if the shorter line is moved farther out to compensate for the fact that it is shorter than its companion. Another interpretation in line with current conceptions of root rectangles would be that the shorter line is moved farther out in order to secure an area (delimited between it and the central line) which is comparable to the area lying between the 10 cm line and the base line.

[6] Moore, H. T., *Op. cit.*, p. 19.

[7] Pierce, E., "Aesthetics of Simple Forms," *Psychol. Rev.*, I, 1894, 483–495 and Vol. 3, 1896, pp. 270 ff.

Pierce took 3 vertical white lines, each one 10 cm in length. Two of them are fixed in vertical position 60 cm apart. The third one is intermediate and is adjustable. Each one of six observers selected a position for the middle line which divided the horizontal distance of 60 cm into the proportions of the golden section! Pierce's interpretation of this result is quite in harmony with our theory of affectivity. The subject does not divide the distance into halves, for this would be too simple and monotonous. Something must be done to give variety. The proportion of the golden section appears to give this without the loss of unity. So far we have considered the results when but three lines are used in which the observer departs from the simple ratio of one to one. But as other lines are added, four and then five, *the tendency to symmetrical arrangement appears* and replaces the preference for the golden section. This means in terms of our theory that the added lines give the complexity which the golden ratio gives; and in order to avoid too much complexity, the subject restores simplicity by simplifying the spacing. This tendency to introduce simple spacing is more marked with five than with four lines. But when six lines are used the observer unexpectedly begins to introduce complexity of spacing. Pierce (rightly) takes this to mean that six lines are so numerous that they are seen by the observer as a whole and not individually. He now seeks to break up this perception of wholeness by introducing *irregularity*. One gets intimations in these simple experiments of a principle which must be coextensive with life. The intellect perhaps would fain have the universe a perfectly simple, symmetrical and unified one. Such a world is one, however, in which our voluntary nature cannot "feel at home"—to use the phrase of James. Just as the intellect desires solutions, our restless conative nature desires "problems." And problem-solving is so inveterately the warp and woof of life that the average observer has an instinctive suspicion of the alleged "reality" of rationalistic

world-views such as those of Plato and Hegel, whose beauty, truth and goodness are "perfect."

Recent experimental work has acquainted us with the so-called law of affective contrast. In the Vassar laboratory[8] observers made absolute judgments as to the pleasantness and unpleasantness of colored papers. Later, 18 colors were preceded in one trial by a color previously selected as pleasant while in another trial they were preceded by a color previously selected as unpleasant. As a result of these experiments and others the writers define a law of affective contrast as follows: "The pleasure of an agreeable experience, and an impression in itself unpleasant may be felt as pleasant if a more unpleasant state has been its antecedent. In like manner unpleasantness may be heightened or even created through contrast with a preceding agreeable affective state." In 1929 this law was verified and stated in more comprehensive form as a result of experiments in the Harvard Laboratory.[9] This more recent formulation holds that the affective value of any experience is a function of *all* of the conscious experiences preceding it, not merely of the event immediately preceding it. The general validity of this law is quite apparent from daily experience. It is nicely illustrated by the onset and recovery from illness. While health is failing, misery is exaggerated by the lingering aura of well-being. The convalescent, on the other hand, finds an unwonted pleasure in mere sight, sound, action, thanks to the lingering aura of disease. Paradoxically, the hedonist's pleasures are made real in part by the pains which he cannot and should not avoid!

The law of affective contrast is in harmony with a conative theory of affection. According to this theory, pleasure and displeasure are results of the facilitations and frustrations of

[8] Washburn, M. F., Bacon, M. M., and Rood, E. A., *American Journal of Psychol.*, 25, 1914, pp. 290–293.

[9] Beebe-Center, J. G., *Am. J. Psychol.*, 41, 1929; Harris, A. J., *Am. J. of Psychol.*, 41, 1929, pp. 617–624.

desire. We need not confine desire to the narrow limits specified by the primal instincts. But, facilitation and frustration are also relative terms: the termination of inhibition has facilitative value. To have some obstruction removed is equivalent to being nearer the goal. In Moore's experiment, for instance, the change from a more to a less dissonant chord would be moving in the direction of harmony and is accepted as such by the feelings of the observer. One can symbolize the situation by the familiar device of positive and negative direction, despite the fact that the numbers remain positive. Troland notes this analogy, and adapts his neurological theory of affection to the law of contrast.[10]

[10] Troland, L. T., *Fundamentals of Human Motivation*, 1928, p. 287.

MENTAL GROWTH—A CRITICISM

G. R. SMITH
University of Toronto

In the article entitled Mental Growth, which appeared in the December, 1930, number of THE JOURNAL OF APPLIED PSYCHOLOGY, (pp. 517–531), the author, A. M. Jordan, regarding the aim of the investigation says, "It is hoped that some light will be thrown upon the general form of the (mental) growth curve; upon the form of curves at different ages, upon the different forms for different levels of brightness at each age, upon the advantages of a group scale in measuring mental growth, and upon the question of the variability of the scores at each age." In his summary of the conclusions regarding these matters which are drawn from the investigations he says, "The mental growth of young children in a community such as we have described in this paper is well represented by a straight line curve which in slope approaches very closely to an angle of 45°." Then he sums up points of evidence for this statement in paragraphs numbered 1, 2, 3, 4, 5, 6.

In this article I am dealing *first* with this "general form" of mental growth curve and with paragraphs 1 and 2 of the evidence, and *secondly* with points 3 and 5 of the evidence.

Since an intelligence test is standardized by making the average score of hundreds of children of a given chronological age, say 12 years, the standard score for that age, it follows that, if a child of CA, 12, makes this standard score for CA, 12, he is at once assigned an MA of 12. Hence when the coordinates used in plotting the mental growth curve of an *ideal average* child, created from the averages of hundreds of real children, are his CA's and his corresponding MA's, they occur in the form—(7, 7), (8, 8), (9, 9)—and the graph

is a perfect straight line with a slope of 45°. Now it would be very strange if the corresponding mental growth curve made by Dr. Jordan *did not* fit this ideal one quite closely since, as he says, there were children of superior, average and inferior mental ability in the group. The drop in the upper end of his curve is due, as he points out, to the facts that he was not testing the same children throughout and that the last group used had a preponderance of children of inferior ability. It thus appears that Dr. Jordan has not determined "the general form of the (mental) growth curve" but has merely shown that this curve *for children of a particular community* fits quite closely the general form of such a curve, which general form has been determined from the beginning.

The evidence submitted in paragraphs 3 and 5 is to the effect that, (1) the gain in mental months between corresponding ages is for the averages of the three groups (high, average, low) the same, and hence the mental growth curves are parallel, (2) the corresponding IQ curves are parallel. The consideration of an ideal situation will throw some light on these matters. Suppose H to be an average high-ability child of CA, 9, and MA, 10; let his gain in CA and MA (using the 45° slope) be 3 years; then at CA, 12, his MA is 13. Similarly suppose A to be an average average-ability child first with a CA and an MA of 9, then 3 years later, with a CA and MA of 12. Also, similarly, suppose L to be an average low-ability child of CA, 9, and MA, 8, at first and then, with 3 years' gain, of CA, 12, and MA, 11. The mental growth curves of these three average children will be parallel straight lines; each will have a slope of 45°. (See Fig. 1.) Note that the mental gain (36 months) of each is the *same* as that of any other. Now consider the IQ's of these three.

$$\text{The IQ of H at CA, 9, is } \frac{10}{9} = 1.111,$$

$$\text{and at CA, 12, is } \frac{13}{12} = 1.083.$$

$$\text{The IQ of A at CA, 9, is } \frac{9}{9} = 1.000,$$

and at CA, 12, is $\frac{12}{12} = 1.000$.

The IQ of L at CA, 9, is $\frac{8}{9} = 0.889$,

and at CA, 12 is $\frac{11}{12} = 0.917$.

For H there is a *decrease* of .028, or 2.5%.
For A " " no *change*.
For L " " an *increase* of .028, or 3.1%.

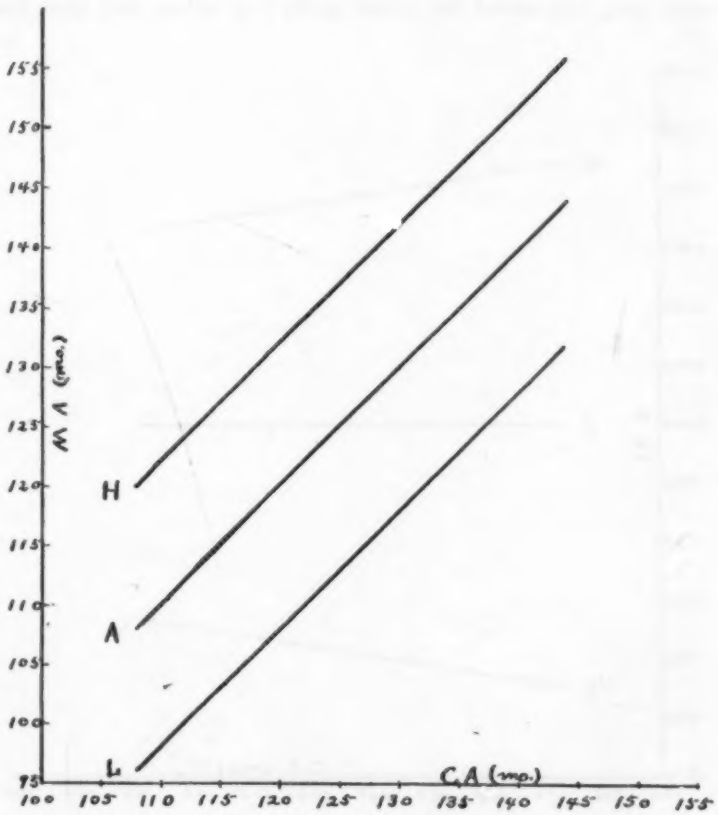


Figure 1.

(A fraction > 1 is decreased, a fraction $= 1$ is unchanged, a fraction < 1 is increased by equal increments to numerator and denominator.) Thus if the MA graphs are parallel straight lines with slopes of 45°, the corresponding IQ graphs can *not* be parallel straight lines; they are converging in form. (See Fig. 2.) Hence the conclusions of paragraphs 3 and 5 are contradictory. Of course the results of the tests can not be expressed in figures as exact as those above, and the author clearly states that the gain in mental months in each case was *about* the same as in any other and that the

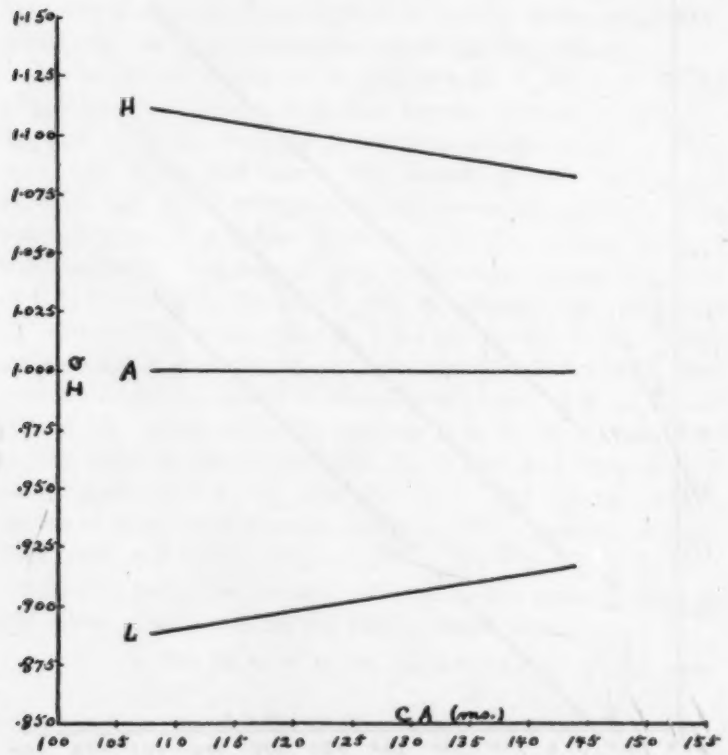


Figure 2.

parallelism of the curves was but a *general tendency*. Yet this again is but a natural conclusion from the form of the curve (straight line, 45° slope) determined from the beginning.

With respect to the matters discussed in this article, the experiment merely shows that these matters do appear in that form in which they were predestined to appear. To get a real mental growth curve we need a measure which can be used without change from infancy to old age.

NOTES AND NEWS

Dr. C. M. Louttit, acting as Assistant Professor of Psychology at Ohio University during the year 1930–31, has been made Director of Clinical Psychology and Assistant Professor in Indiana University. He is the successor of Dr. H. H. Young, who with Mrs. Young had so ably carried on work in Clinical Psychology both at Indiana University and at Indianapolis that these centers had come to be recognized as representatives of the best to be found in this rapidly growing field.

A review of educational legislation enacted in the forty-eight states during the past two years reveals that major educational problems in the United States include school finance, teacher certification and pensions, school attendance, curriculum changes, state and county administration, school consolidation, pupil transportation, and education of physically and mentally handicapped children, according to the Office of Education, U. S. Department of the Interior.

On April 25 and 26, 1931, Antioch College, Yellow Springs, Ohio, held a Symposium on the question "What Is Life For?" The following well-known speakers discussed the meaning of life for the individual from the point of view of Judaism, Christianity and Humanism: James G. Heller, Rabbi of the Plum Street Temple of Cincinnati; Rev. Abbott B. Cohoe, Pastor of the First Baptist Church of Montclair, N. J.; and Dr. A. Eustace Haydon, Professor of Comparative Religion, University of Chicago. Each address was followed by round-table discussions for all who were interested. The final session was planned as a "conversation" between the three speakers in the presence of the entire gathering, with President Arthur E. Morgan, of Antioch College, presiding.

Dr. S. M. Stoke, for four years a member of the psychological staff of Ohio University, has been made head of the Department of Education in Mt. Holyoke College for the coming year. His rank is that of Associate Professor and he will offer courses in Educational Psychology, Educational Measurements and Vocational Guidance.

Figures presented in "Per Capita Costs in City Schools," a biennial study prepared by the Office of Education, Department of the Interior, show that sixty cents buys one day's education for a child in the average

city public school. Statistics from a representative group of American cities show that the total average annual cost per city child for teachers' services for nine months, for heat, janitor service and school supplies, is around $108.87.

Dr. L. L. Thurstone and Charles K. A. Wang have included in a series of scales on social attitudes one entitled "Attitude Toward Birth Control." This appears as Scale No. 21, Forms A and B. Each form contains twenty questions expressing different attitudes toward birth control. In filling out the blank one is instructed to place a check-mark before the statement if he agrees, a double check-mark if he strongly agrees, and a cross if he disagrees with the statement. Tables are worked out in which each statement is given a certain value. These Scales may be obtained from the University of Chicago Press.

A recent survey by the Federal Office of Education reveals the following interesting facts concerning the training of teachers of the 153,000 one-room rural schools of the United States: If all the teachers of these schools stood side by side, their ranks would extend in an unbroken line 87.1 miles. Assuming that the teacher with the least amount of training stood at one end and the one having the largest amount of training at the other, it would be necessary to walk a distance of 8½ miles before coming to a teacher with a training equivalent to two years of high school. One would have to walk half the entire distance to reach the teacher with training equal to high school graduation; 67⅝ miles to reach the first teacher with the equivalent of two years of normal school education; and 74.1 miles to reach the one with a college education.

The composite teacher of this group would be a woman about twenty-seven years old with a total education of four years and one month above the grade school; her teaching experience would total 2 years and 6 months; her salary would be $874 per year with a total of 152 days; and she would have under her care a total of 22 farm children.

The Fox Film Corporation announces that the District of Columbia and all the states of the Union but two have accepted the invitation of President Hoover to delegate a boy, a girl and a school official to meet in Washington on July 6 to participate in a nation-wide demonstration of the value of talking motion pictures in our public schools. The test will be conducted by a committee of educational experts appointed by the United States Commissioner of Education. The chairman of the committee is Dean William C. Ruediger of the School of Education of George Washington University. After one morning devoted to an examination in certain phases of geography, science, American history and

civics, the children will be examined daily for three days after being given Visual Education in these subjects by the use of sound motion pictures, to discover what they have added to their knowledge. It is believed that the support given to Visual Education by President Hoover will lend great impetus to a movement that has been in much favor with educators throughout the country for several years and with closer cooperation between producers and educators the use of sound motion pictures in the schools will be greatly extended.

Dr. Paul F. Farnsworth has reported the following news notes from Stanford University:

Dr. H. D. Carter has been reappointed a Social Science Research Council Fellow, and Dr. P. E. Fields, a National Research Council Fellow. Both will continue their work in psychology at Stanford.

Mrs. R. H. Thomson becomes Assistant Professor of Education and Psychology at San Francisco State Teachers College.

Mr. C. R. Carpenter has been appointed a National Research Council Fellow. He will do his work at Yale and at the Anthropoid Experiment Station at Orange Park, Florida.

Mr. R. G. Bernreuter becomes Assistant Professor in the Extension Division at Pennsylvania State College.

Dr. William D. Commins has been appointed Instructor of Psychology at the Catholic University of America.

Dr. Warner Brown of the University of California was on the 1931 summer staff of the Psychology Department, at Stanford.

Teachers College, Columbia University, presented Dr. Egon S. Pearson, Assistant Editor of Biometrika and Senior Lecturer in the Biometric Laboratory, University College, University of London, in a series of lectures on "Trends in Statistical Research," on May 1, 4, 6 and 7, 1931. No doubt the chief interest of readers of the Journal in this note will be because this marks the opening of the Teachers College Statistical Laboratory.

BOOK REVIEWS

DANIEL KATZ & FLOYD H. ALLPORT. *Students' Attitudes.* A Report of the Syracuse University Reaction Study. Syracuse, N. Y.: The Craftsman Press. 1931. 408 pp.

Attitudes, among all the problems of modern social psychology, hold a place of unique importance. From the point of view of psychology they involve questions of emotions, sentiments and other habitual ways of behaving, while to the sociologist they contribute to the understanding of groups and institutions as well as to the understanding of social interaction. The gathering of data on such problems necessitate the use of a methodology not yet adequately explored and therefore requiring a careful and critical interpretation of the results. In spite of the possible pitfalls the present authors have produced a notable piece of work on the existing attitudes of college students.

The data upon which the discussion rests were secured from a questionnaire answered by 4248 students at Syracuse University on May 10, 1926. The questionnaire itself was notable in that it was built on the reports of actual student interest as shown in conversation and discussion as observed by a student committee. Rapport with the student body was established through campus newspaper and other publicity, although the types of question were not revealed. The great care shown in all the preliminaries to the collection of data is in itself almost sufficient to establish the quality of this study from the experimental point of view.

The questionnaire contained seventy-five items, some of which had two or more sub-divisions. It would be impracticable to attempt a summary of the whole data here. Some idea of the scope of student interest covered is indicated by the following topics: reasons for coming to college; college activities; attitudes toward studies; personal ideals; adjustment to the college situation; adjustment to the problem of choosing a vocation; fraternity vs. non-fraternity attitudes; snobbishness; cribbing; sex morals and co-education; religious life.

Using his prerogatives the reviewer shall mention only a few of the conclusions which seem important to him. The practical value of a college degree received more emphasis than the cultural advantages although scholarly interests attained "somewhat greater emphasis in the later years of college life." Another indication that seems important is shown in this quotation: "Our results suggest that the stability of purpose and self-direction given by certainty concerning one's future voca-

tion is an important basis not only for the selection of a curriculum and the serious interest which a student feels in his work, but for the development of his latent abilities and his entire personality.''

A third finding of importance is in relation to religion. In spite of the rather wide-spread belief that college students are irreligious or heretical this study found 64 per cent. taking a strictly orthodox view of the nature of God, while only 5 per cent. indicated atheism. Although these three conclusions are only a few of the many interesting suggestions the study makes they are all the available space will allow mention of.

Two supplementary chapters discuss the techniques of attitude measurement and a survey of college attitude studies. A bibliography of 82 titles accompany these chapters.

JOHANNES LINDWORSKY. *Experimental Psychology.* (Translated by H. R. DeSilva) New York: Macmillan Co. 1931 + xix 406 pp. $3.75.

Adelbert Ford. *Group Experiments in Elementary Psychology.* New York: Macmillan Co. 1931. vi + 241 pp.

These two works appearing at the same time with the same imprint afford a study in contrasts. Apart from the fact that the first is a textbook, while the second is a laboratory manual they represent opposing points of view, or at least differences in emphasis.

Lindworsky says, ''All of these phenomena, which, by their nature, are immanently conscious and immediately apprehensible only for the experiencing subject, form, therefore, the subject matter of experimental psychology.'' This is in nice contrast to the definition of Ford who says, ''From the standpoint of the experimental laboratory, we shall not go far astray if we define psychology as the study of human and animal behavior, with its attributes, variations, and determinants.''

The German author stays consistently with his point of view discussing in turn psychical elements, i.e., sensations, images, feelings and elementary volitions; revival of images and association; higher mental achievements including cognition, higher feelings and volition; socially influenced mental achievements, as language, morals, art and religion; and finally devotes Book Five to exceptional mental states, i.e., sleep, dreams and hypnosis. Throughout the discussion the author consistently maintains the introspectionist point of view and is only incidentally concerned with behavior. The account is convincingly written and has been ably translated. It is pertinent to note that no English or American, and very little French literature is quoted.

Ford's manual keeps as consistently as possible to the definition quoted above. He avoids the method of introspection except in some experiments on the senses where it is after all the most economical one to use.

Four experiments are devoted to statistical methods, fifteen to group demonstrations, sixteen to group experiments in general psychology, and nine to group demonstration experiments in applied psychology. The whole field of sensory phenomena, reflexes, learning, attention, perception, intelligence, mental tests, etc., is sampled. For each experiment the purpose, history and orientation, procedure and references are given, together with well selected questions to bring out the significance of the results.

Having given brief pictures of these two books it remains to evaluate them in terms of classroom use. The Macmillan Company is to be commended for publishing Lindworsky's book in present day America. The present reviewer predicts that this work will be avoided as heretical, scorned as unscientific, or stacked on library shelves somewhat as a curiosity. He also hopes that his prophecy is wrong because this book will serve as an excellent antidote to too much Behaviorism, even though it does not succeed in winning converts to itself. It is recommended that every graduate student in the majority of psychology departments in the United State read this book to secure an idea of a point of view all too evidently absent from those departments.

The manual of group experiments is to be highly recommended to all psychology departments whose equipment does not allow individual experimentation in the elementary course. Its organization and presentation as well as the physical make-up of the book make it extremely well adapted to the elementary student.

WATSON, JOHN B. *Behaviorism.* New York: W. W. Norton & Co., 1930. xi + 308 pp.

Dr. Watson is a most stimulating and sane author when he writes of Behaviorism; he is most irritating and almost fanatical when he labors with "psychology." One wonders if sometimes his writing is not defense (psychoanalytic word to be sure, and yet not anti-behavioristic). Must there be the quarrel which he seems to insist upon between psychology and its young offspring behaviorism?

Introspectional psychology (using the term as the name of a school, not a method) was, and is, interested in certain phenomena of nature—sensations, images, perceptions, in a word those phenomena which may be termed experiential. History amply demonstrates that this school became ensnared in meshes of its own making. The workers either lost sight of the equally important problems (from a practical point of view perhaps more important) of behavior, or in a few instances tried to explain them in introspectional terms. The sterility of their results—in dealing with behavior—would seem to be evident.

But let us look at the other side of the picture. Is not the behaviorist falling into the same evil ways? Does he not deal with behavior to the

exclusion of other phenomena? If called upon to give an account of experience does he not step aside, or else attempt a behavioristic account which may or may not be adequate?

The truth, as is so often the case, probably lies between the two extremes. However, there seems to be no need to add more words to the controversy. Present trends indicate that its resolution is not far distant.

Regarding the book itself little need be said. In physical format and in tone of writing this edition is much superior to that of 1925. The author says that the overstatements and exaggerations of the first edition have been remedied. It is evident that new material has been added and the experimental evidence is up to date when we find reference to a paper published in September 1930 while the preface is dated August 1930! As with the previous edition the work is addressed to the intelligent layman, although, except for the omission of many topics often considered important, there is no objection to its use as a text.

C. M. Louttit,
Ohio University.

Earl D. Bond, M.D., and Kenneth E. Appel, M.D. *The Treatment of Behavior Disorders Following Encephalitis: An Experiment in Re-Education.* The Commonwealth Fund Division of Publications, New York City, 1931. 163 pp. Price $1.75.

In this short volume the authors give an account of experiments in re-educating children who were suffering from behavior disorders following encephalitis. The experiments took place in the Pennsylvania Hospital. Forty-eight such children were involved and also a group of fourteen problem children who served as controls. The period covered by the report extends from the founding of the class in December, 1924 to January, 1930.

A considerable amount of space is devoted to the description of the children, their peculiarities, and the training given them. The experiences of the hospital and their general conclusion can best be gained from their own words: "When the Pennsylvania Hospital began these studies, the staff was thinking in terms of X-ray examinations and treatment, chemistry, and neuropathology, and now after four years it is thinking in terms of education and mental hygiene." (Pp. 147–148.) All but two of the forty-eight children showed improvement while in the hospital. Apparently it is much easier for a small number of trained attendants to care for a considerable number of such children under proper conditions than for parents to care for one. Routine often produced a surprising influence for good. A number of children were restored to normal sleeping habits after a short stay in the hospital

school with scarcely any effort on the part of the attendants. The records of those discharged from the hospital are not so promising. Twenty-six children were discharged but only seven made satisfactory progress afterwards. Seven others were doubtful, six were feebleminded and little could be expected of them, six showed definite retrogression. It must be remembered that one hundred per cent. of these children were retrograding when taken into the school, and that a number were sent back to homes which were definitely a bad influence. When these facts are taken into consideration, the number of restored children does not seem so small. From occasional glimpses into the home backgrounds of some of these children, one is caused to wonder whether a considerable amount of the behavior disorders was not due to the impact of an unwise or bad home upon a long convalescent and little understood child, rather than to the disease itself.

Those interested in problem children will find the book interesting, candid, and decidedly helpful. It is a worthy addition to an inadequate literature on the subject.

BURKS, BARBARA S., JENSEN, DORTHA W., TERMAN, LEWIS M. *Genetic Studies of Genius, Vol. III. The Promise of Youth.* Stanford University Press, 1930. Pp. xiv and 508. $6.00.

This volume is concerned with a follow-up of the original gifted children reported on in volume I. Not all of them were available for further study, but only a surprisingly few were lost. The second study did not include a physical examination but did do extensive work in the field of mental and educational status and achievement. A considerable number of case studies are reported and especial attention was given to the literary expressions of some members of the group. Other exceptional talents in music, science and art were given somewhat briefer notice.

The most outstanding fact is that the children, as a whole, show the same superiorities in 1927–28 that they did in 1921–22. For a number of individuals this is untrue. Girls, especially, showed a tendency to lower their IQ's in adolescence. It is to be regretted that no physical examination of the group was made in order to determine whether this was due to differing rates of maturation. The proportion of gifted boys to gifted girls seems to increase as maturity approaches. Aside from this fact there seems to be no marked change in abilities. Intellectual interests predominate in the lives of these children, now as formerly. Their achievements in school have for the most part been superior. There are cases, however, of indifference, bad character traits, unhappy homes, wrong guidance, etc., which demonstrate that the gifted child is by no means certain even of finishing school with credit. The bulk of these

children have the professions as their vocational objective. A few have already gone to work and some have married. The latter, to the authors' disappointment, have married people inferior to themselves. Measures of masculinity and femininity reveal that the boys are normally masculine but that the girls are considerably more masculine in their interests than the average girl.

Case studies of development, socially as well as intellectually, are given. These reveal that the gifted child may mature physically and socially at such a rate as to cause no maladjustments when double promotions are given in school. However, each case must be considered on its own merits for some are reported in which educational and mental acceleration were not accompanied by a corresponding increment of physical growth and social poise. One such case appears to be leading toward a career of research. Another gives evidence of leading to nowhere except bewildered incompetence. Evidently acceleration in school as a blanket policy is untenable.

A considerable amount of space is devoted to the development of a literary scale for judging the merits of literary productions of juveniles. This scale is applied to the literary efforts of a number of gifted children and to those of several authors of note. The comparisons are not at all unfavorable to the juvenilia of the study group. However, the authors refrain from hailing their prodigies as prospective Brownings, Shelleys, etc.

There are many more interesting facts which the reviewer's space does not permit him to mention, but which will justify the reader's attention. The tone of the volume is conservative and judicial rather than enthusiastic and propagandistic. There is evidence of much thought as well as the turning of the wheels of the research techniques. If the two remaining studies of these gifted children to be made at much later dates can be made in the same scientific spirit, the educational and psychological world will be vastly richer. We have long been in need of extensive and intensive genetic studies as well as the common "cross section" study of life. These studies attempt to supply that need.

STUART M. STOKE,
Ohio University.

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Ohio University, Athens, Ohio.

Animal Drive and the Learning Process. EDWIN B. HOLT. Henry Holt and Company, New York, 1931. 307 pp.

Certain Relationships between Scholarship in High School and in College. LENA JAMES HAWKS. Johns Hopkins University Studies in Education, No. 15. Johns Hopkins Press, Baltimore, 1931. 58 pp.

Child Psychology. JOHN J. B. MORGAN. Richard R. Smith, New York City. $3.00. 474 pp.

Current Practice in the Construction of State Courses of Study. MINA M. LANGVICK. U. S. Department of Interior, Bulletin No. 4, 1931. Government Printing Office, Washington, D. C. 15 cents. 84 pp.

The Dissatisfied Worker. V. E. FISHER AND JOSEPH V. HANNA. Macmillan Company, 1931. $2.50. 260 pp.

L'Education Fonctionnelle. EDWARD CLAPAREDE. Delachaux and Niestle, S. A. 26 Rue St-Dominique, Paris. 263 pp.

Experimental Social Psychology. GARDNER MURPHY AND LOIS BARCLAY MURPHY. Harper and Brothers, New York City, 1931. $3.50. 709 pp.

Experimental Psychology. JOHANNES LINDWORSKY. Macmillan Company, New York, 1931. 406 pp.

Fields of Psychology. GLENN DEVERE HIGGINSON. Henry Holt and Company, New York, 1931. $2.75. 613 pp.

Genius and Creative Intelligence. N. D. M. HIRSCH. Sci-Art Publishers, Harvard Square, Cambridge, Mass., 1931. $4.50. 339 pp.

The Influence of Training on Changes in Variability in Achievement. HOMER B. REED. Psychological Review Company, Vol. XLI, No. 2, 1931. Princeton, N. J. 59 pp.

Introduction Generale Psychologie. Manuel de Philosophie, Vol. I. LIBRARIE ARMAND COLEN, 103 Boulevard Saint-Michel, Paris. 722 pp.

Order of Birth, Parent-Age, and Intelligence. L. L. THURSTONE AND RICHARD L. JENKINS. University of Chicago Press, 1931. Chicago, Ill., $3.00. 135 pp.

Psychological Principles in Automotive Driving. ALBERT P. WEISS AND ALVHH R. LAUER. Contributions in Psychology, No. 11, Ohio State University Studies, 1931. Columbus, Ohio. $1.50. 165 pp.

The Psychology of Exceptional Children. NORMA V. SCHEIDEMANN. Houghton Mifflin Company, Boston, 1931. $3.25. 520 pp.

La Science du Caractere. W. BOVEN. Delachaux et Niestle, S. A. Neuchatel and Paris, 1931. 4 Rue d l'Hospital, 1931. 351 pp.

The Treatment of Behavior Disorders Following Encephalitis. EARL R. BOND AND KENNETH E. APPEL. Commonwealth Fund, Division of Publication, 41 E. 57th Street, New York. $1.75. 163 pp.

* Mention here does not preclude further comment.

Directory of American Psychological Periodicals

(Continued)

Psychological Abstracts—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6.00. 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.: Clark University Press. Subscription $7. 500–600 pages annually. Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.: Clark University Press. Subscription $7.00. 500–600 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—Baltimore: The Williams & Wilkins Co. Subscription $5.00. 500 pages, approximately. Edited by Walter V. Bingham. Bi-Monthly.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Philosophy—New York: 515 West 116th Street. Subscription $4.00. 728 pages per volume. Founded 1904. Bi-weekly. Edited by F. J. E. Woodbridge, Wendell T. Bush, and H. W. Schneider.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription, $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Directory of American Psychological Periodicals

American Journal of Psychology—Ithaca, N. Y.: Cornell University. Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, E. M. Dallenbach, M. Bentley, E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology and The Pedagogical Seminary—Worcester, Mass.: Clark University Press. Subscription $7. 600 pages annually. Edited by Carl Murchison. Quarterly. Child behavior, differential and genetic psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company. Subscription $5.50. 540 pages annually. Bi-monthly. General Psychology. Edited by Howard C. Warren. Founded 1894.

Psychological Monographs—Princeton, N. J.: Psychological Review Company. Subscription $6.00 per vol. 500 pp. Ed. by Raymond Dodge. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.: Psychological Review Company. Subscription $4.00. 300–400 pp. Edited by W. S. Hunter. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.: Psychological Review Company. Subscription $6.00. 720 pages annually. Psychological literatrre. Monthly. Edited by Edward S. Robinson. Founded 1904.

Archives of Psychology—New York: Columbia University. Subscription $6. 500 pp. ann. Ed. by R. S. Woodworth. Published without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J. American Psychological Association. Subscription $6. Foreign, $5.25. Editor, Henry T. Moore. Quarterly. 448 pages annually. Abnormal and Social. Founded 1906.

Psychological Clinic—Philadelphia: Psychological Clinic Press. Subscription $3.00. 288 pages. Ed. by Lightner Witmer. Without fixed dates (9 Numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.: 3617 10th St., N. W. Subscription $6. 500 pages annually. Psychoanalysis. Quarterly. Ed. by W. A. White and S. E. Jelliffe. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J. Psychological Review Company. 500 pages annually. Experimental. Subscription $6.00. Bi-monthly. Edited by C. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Lancaster: The Science Press Printing Co. Subscription $3.00. 400 pages annually. Bi-monthly. Edited by James P. Porter, Ohio University, Athens, Ohio. Founded 1916.

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research.

Genetic Psychology Monographs—Worcester, Mass.: Clark University Press. Subscription $7.00 per volume. Two volumes per year, 600 pages each. Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

(*Continued on Third Cover*)